

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 13:46:46 ; Search time 5930.9 Seconds
(without alignments)
39.846 Million cell updates/sec

Title: US-09-396-196F-8
Perfect score: 25
Sequence: 1 gcattgcattcctcgtcagtgacg 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1022815 segs, 4726426750 residues
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estin6:*
69: em_estin7:*
70: em_estin8:*
71: em_estin9:*
72: em_estin10:*
73: em_estin11:*
74: em_estin12:*
75: em_estin13:*
76: em_estin14:*
77: em_estin15:*
78: em_estin16:*
79: em_estin17:*
80: em_estin18:*
81: em_estin19:*
82: em_estin20:*
83: em_estin21:*
84: em_estin22:*
85: em_estin23:*
86: em_estin24:*
87: em_estin25:*
88: em_estin26:*
89: em_estin27:*
90: em_estin28:*
91: em_estin29:*
92: em_estin30:*
93: em_estin31:*
94: em_estin32:*
95: em_estin33:*
96: em_estin34:*
97: em_estin35:*
98: em_estin36:*
99: em_estin37:*
100: em_estin38:*
101: em_estin39:*
102: em_estin40:*
103: em_estin41:*
104: em_estin42:*
105: em_estin43:*
106: em_estin44:*
107: em_estin45:*
108: em_estin46:*
109: em_estin47:*
110: em_estin48:*
111: em_estin49:*
112: em_estin50:*
113: em_estin51:*
114: em_estin52:*
115: em_estin53:*
116: em_estin54:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.2	76.8	265	163	BE125672 DG1_54.F0
2	19.2	76.8	353	114	AM287736 LG1_271.G
3	19.2	76.8	542	166	BE357782 DG1_22.F0
4	18.2	72.8	372	189	W00149 T9ESTzy77e0
5	18.2	72.8	431	122	AW925187 WSL_76.E0
6	18.2	72.8	544	239	AE143198 SP_0028.B
7	18.2	72.8	613	228	AE143198 SP_0028.B
8	17.8	71.2	538	30	AV439422 HS_5083.B
9	17.6	70.4	508	227	AV439422 HS_5083.B
10	17.6	70.4	574	167	AQ358636 HS_5030.A
11	17.6	70.4	576	163	BE405752 WHE1214_H
12	17.6	70.4	1032	168	BE131409 L48-1384T
13	17.2	68.8	325	137	BE682210 602117637
14	17.2	68.8	343	158	BE595789 P11_54.A0
15	17.2	68.8	349	150	BE604656 CM0-CT031
16	17.2	68.8	364	218	BE524069 UI-R-C0-h
17	17.2	68.8	654	14	AF116319 AF116319
18	17.2	68.8	719	122	AA946503 EST20002
19	17.2	68.8	988	222	AA946503 EST20002
20	17.2	68.8	997	222	AA946503 EST20002
21	17.2	68.8	1081	222	AA946503 EST20002
22	17.2	68.8	1095	221	AA946503 EST20002
23	17.2	68.8	113	7	AA422533 VF14A04.S
24	17.2	68.8	300	148	BF386481 UI-R-CA1-
25	17.2	68.8	332	25	AT724043 RH121_25
26	17.2	68.8	360	250	AT724043 RH121_25
27	17.2	68.8	421	120	AM747232 WSL_66.E0
28	17.2	68.8	432	148	BE387098 UI-R-CA1-
29	17.2	68.8	446	175	BE387098 UI-R-CA1-
30	17.2	68.8	456	18	BE387098 UI-R-CA1-
31	17.2	68.8	463	175	BE387098 UI-R-CA1-
32	17.2	68.8	471	173	BE387098 UI-R-CA1-
33	17.2	68.8	473	173	BE387098 UI-R-CA1-
34	17.2	68.8	476	148	BE387098 UI-R-CA1-
35	17.2	68.8	489	122	BE387098 UI-R-CA1-
36	17.2	68.8	497	166	BE387098 UI-R-CA1-
37	17.2	68.8	502	172	BE387098 UI-R-CA1-
38	17.2	68.8	506	244	BE387098 UI-R-CA1-
39	17.2	68.8	510	148	BE387098 UI-R-CA1-
40	17.2	68.8	511	166	BE387098 UI-R-CA1-
41	17.2	68.8	511	166	BE387098 UI-R-CA1-
42	17.2	68.8	515	235	BE387098 UI-R-CA1-
43	17.2	68.8	518	166	BE387098 UI-R-CA1-
44	17.2	68.8	519	166	BE387098 UI-R-CA1-
45	17.2	68.8	529	258	BE387098 UI-R-CA1-

ALIGNMENTS

RESULT 1	BE125672	265 bp	mRNA	EST	19-JUL-2000
LOCUS	DG1_54.E04.b1_A002	Dark Grown 1 (DG1)	Sorghum bicolor cDNA, mRNA		
DEFINITION	sequence.				
ACCESSION	BE125672				
VERSION	BE125672.1	GI:6548262			
KEYWORDS	EST.				
SOURCE	Sorghum.				
ORGANISM	Sorghum bicolor				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.				
AUTHORS	1 (bases 1 to 265)				
TITLE	Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and Pratt, L.H.				
COMMENT	An EST database from Sorghum: dark-grown seedlings				

JOURNAL COMMENT

Unpublished (2000)
Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu

FEATURES

Source
Location/Qualifiers
1..265
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DG1)"
/note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."
BASE COUNT
54 a 98 c 75 g 38 t
Query Match
Best Local Similarity 76.8%; Score 19.2; DB 163; Length 265;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY

2 cattgcatctcgtcagtcag 25
||||| ||||| ||||| |||||
Db 61 cattgcatctcgtcagtcag 84

RESULT 2

LOCUS AM287736 353 bp mRNA EST 19-JUL-2000

DEFINITION

LG1_271.G07.b1_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA

ACCESSION

AM287736 353 bp mRNA EST 19-JUL-2000

VERSION

AM287736.2 GI:6859727

KEYWORDS

EST.

SOURCE

Sorghum.

ORGANISM

Sorghum bicolor

REFERENCE

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.

AUTHORS

1 (bases 1 to 353)

TITLES

An EST database from Sorghum: light-grown seedlings

COMMENT

Unpublished (2000)
On Jan 6, 2000 this sequence version replaced gi:6677580.
Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude polyA, vector and regions below phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 352
POLYA=No.

FEATURES

Location/Qualifiers
1..353
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Light Grown 1 (LG1)"
/note="Organ: 10- to 14-day-old light-grown (greenhouse)

REFERENCE
1 (bases 1 to 265)
Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and Pratt, L.H.

TITLE
An EST database from Sorghum: dark-grown seedlings

seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI
; The library was made from poly-A RNA in the cloning
vector Lambda Zap II. Clones to be sequenced were
prepared by mass excision."

BASE COUNT 55 a 120 c 123 g 55 t

ORIGIN

Query Match 76.8%; Score 19.2; DB 114; Length 353;
Best Local Similarity 87.5%; Pred. No. 67;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cattgacatctgcagtgacg 25
||||| ||||| ||||| |||||

Db 47 CATTGCCCTCTCTGTCAGCTGCAG 70

RESULT 3

BE357782

LOCUS

DEFINITION DGL_22.F01.b1_A002 Dark Grown 1 (DGI) Sorghum bicolor cDNA, mRNA

ACCESSION

BE357782

KEYWORDS

EST.

SOURCE

ORGANISM

Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoidae; Andropogoneae; Sorghum.

1 (bases 1 to 542)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

An EST database from Sorghum: dark-grown seedlings

Unpublished (2000)

Contact: Cordonnier-Pratt MM

Department of Botany

The University of Georgia

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 542 1805

Email: mmp@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions

below Phred quality 16. The threshold for highest quality sequence

is 20.

Seq primer: JEN REV

High quality sequence stop: 537

POLYA=No.

FEATURES

SOURCE

Location/Qualifiers

1..542

/organism="Sorghum bicolor"

/db_xref="taxon:4558"

/clone_lib="Dark Grown 1 (DGI)"

/note="Organ: 5-day-old dark-grown seedlings; Vector:

Lambda Zap; Site_1: XhoI; Site_2: EcoRI. The library was

made from poly-A RNA in the cloning vector Lambda Zap II.

Clones to be sequenced were prepared by mass excision."

BASE COUNT 80 a 196 c 187 g 79 t

ORIGIN

Query Match 76.8%; Score 19.2; DB 166; Length 542;

Best Local Similarity 87.5%; Pred. No. 70;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cattgacatctgcagtgacg 25

||||| ||||| ||||| |||||

Db 61 CATTGCCCTCTCTGTCAGCTGCAG 84

RESULT 4

W00149/c

LOCUS

372 bp mRNA EST 22-MAY-2000

DEFINITION T9ESTy77e07.r1 T9RH Tachyzoite cDNA Toxoplasma gondii cDNA clone

ACCESSION

W00149

VERSION

W00149.1 GI:1271572

KEYWORDS

EST.

SOURCE

ORGANISM

Toxoplasma gondii

Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;

Sarcocystidae; Toxoplasma.

1 (bases 1 to 372)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: toxo@wustl.edu

David Sibley at toxo@wustl.edu for further information

relating to organism, clone or library availability.

Seq primer: T3

High quality sequence stop: 321.

LOCATION/Qualifiers

1..372

/organism="Toxoplasma gondii"

/strain="RH"

/db_xref="taxon:5811"

/clone_lib="T9RH Tachyzoite cDNA"

/lab_host="X11-Blue MRF"

/note="Vector: Lambda Zap; Site_1: EcoRI; Site_2: XhoI;

Toxoplasma RH strain tachyzoites were grown in human

foreskin fibroblast cultures in vitro. The library was

constructed by K.L. Wan, Cambridge University. cDNAs were

synthesized from polyA RNAs by oligo d(T) priming and

directionally cloned into the EcoRI to XhoI sites of the

Lambda ZapII vector using the Zap-cDNA synthesis kit

(Stratagene). WARNING: the library contains a small

percentage of cDNAs derived from the human host cells."

BASE COUNT 103 a 111 c 85 g 72 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 189; Length 372;

Best Local Similarity 87.0%; Pred. No. 2e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 attgacatctgcagtgacg 25

||||| ||||| ||||| |||||

Db 208 ATTTCGTCATCGTCGTCGTCAG 186

RESULT 5

AW925187/c

LOCUS

DEFINITION WS1_76_E02.b1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,

mRNA sequence.

ACCESSION

AW925187

VERSION

AW925187.1 GI:8091013

KEYWORDS

EST.

SOURCE

ORGANISM

Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoidae; Andropogoneae; Sorghum.

1 (bases 1 to 431)

REFERENCE

AUTHORS

TITLE

An EST database from Sorghum: water-stressed plants

Query Match 72.8%; Score 18.2; DB 228; Length 613;
 Best Local Similarity 87.0%; Pred. No. 2.1e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 attgcagtcctgcagtcagcag 25
 ||||| ||||| ||||| ||||| |||||

Db 158 ATTTCTATCTCTCTCATGTGCAG 180

RESULT 8
 AVA39422 538 bp mRNA EST 23-AUG-2000
 LOCUS AVA39422 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
 DEFINITION PS050C10.1 r 5', mRNA sequence.
 AVA39422
 AVA39422
 AVA39422.1 GI:8594647
 EST.
 SOURCE Porphyra yezoensis.
 ORGANISM Porphyra yezoensis.
 Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
 Porphyra.
 1 (bases 1 to 538)
 Nikaido, I., Asamizu, E., Nakajima, M., Nakamura, Y., Saga, N. and
 Tabata, S.
 Generation of 10,154 expressed sequence tags from a leafy
 gametophyte of a marine red alga, Porphyra yezoensis
 JOURNAL DNA Res. 7, 223-227 (2000)
 MEDLINE 20363100
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yakuza 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
 Source
 1..538
 /organism="Porphyra yezoensis"
 /strain="TU-1"
 /db_xref="taxon:2788"
 /clone="PS050C10.1"
 /clone_lib="Porphyra yezoensis TU-1"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
 XhoI"

BASE COUNT 114 a 158 c 170 g 96 t

ORIGIN

Query Match 71.2%; Score 17.8; DB 30; Length 538;
 Best Local Similarity 90.5%; Pred. No. 3.3e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gcaattcgatcctgcagtcagc 21
 ||||| ||||| ||||| ||||| |||||

Db 371 GCATTTCATCCCGCGTCAGGT 351

RESULT 9
 AQ358636 508 bp DNA GSS 06-MAR-1999
 LOCUS AQ358636 HS.5030.AL.C01.T7 RPII Human Male BAC Library Homo sapiens
 DEFINITION genomic clone Plate=606 COL-1 Row=E. DNA sequence.
 AQ358636
 AQ358636
 AQ358636.1 GI:4207512
 GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 508)
 Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
 Kehler, A., Shaker, R., Furlong, J., Young, D., Zhao, S., Adams, M.D. and
 Hood, L.
 Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome

REFERENCE
 AUTHORS

TITLE

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 606 Row: E Column: 1
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 508.
 Location/Qualifiers
 1..508
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=606 COL-1 Row=E"
 /clone_lib="RP11 Human Male BAC Library"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBAC3.6; RPII Human Male BAC Library"

BASE COUNT 147 a 97 c 95 g 153 t 16 others

ORIGIN

Query Match 70.4%; Score 17.6; DB 227; Length 508;
 Best Local Similarity 80.0%; Pred. No. 4e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gcaattcgatcctgcagtcagc 25
 ||||| ||||| ||||| ||||| |||||

Db 98 GCANAACTATCCGCGTCAGTGCAG 74

RESULT 10
 BE405752 574 bp mRNA EST 21-JUL-2000
 LOCUS WHE1214_H08_016ZS wheat etiolated seedling root cDNA library
 DEFINITION Triticum aestivum cDNA clone WHE1214_H08_016, mRNA sequence.
 BE405752
 BE405752
 BE405752.1 GI:9365220
 EST.
 SOURCE bread wheat.
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 1 (bases 1 to 574)
 Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han
 , P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J.,
 Seaton, C.L. and Tong, J.C.
 The structure and function of the expressed portion of the wheat
 genomes
 UNPUBLISHED (2000)
 Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: oanderson@pw.usda.gov
 Sequence have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: Strategene SK primer.
 Location/Qualifiers
 1..574
 /organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone="WHE1214_H08_016"

FEATURES
 source

```

/clone_lib="Wheat etiolated seedling root cDNA library"
/tissue_type="Root"
/dev_stage="Five day old etiolated seedling"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-Zap XR, excised phagemid;
Site_1: EcoRI; Site_2: XhoI; Seeds were surface-sterilized,
germinated and grown aseptically in the dark at room
temperature on filter paper with water, nystatin and
cefotaxime in covered crystallization dishes. Roots were
harvested. The tissue, total RNA, and poly(A) RNA were
prepared. A cDNA library was made, and the cDNA clones
were in vivo excised to give Bluescript phagemids in the
TJ Close Lab (Choi, Close, Fenton) at the University of
California, Riverside. Plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."
BASE COUNT      155 a      125 c      153 g      141 t
ORIGIN

Query Match      70.4%; Score 17.6; DB 167; Length 574;
Best Local Similarity 83.3%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gattcgatcctgcagtgagc 24
||||||| 1 1111111 111
Db 289 GCATTTCGTCGCTCGTCGAGGCA 312

RESULT 11
BE31409 576 bp mRNA EST 20-FEB-2001
LOCUS L48-138473 Ice plant Lambda Uni-Zap XR expression library, 48 hours
DEFINITION NacI treatment Mesembryanthemum crystallinum cDNA clone L48-1384,
mRNA sequence.
ACCESSION BE31409
VERSION BE31409.1 GI:8578772
KEYWORDS common ice plant.
SOURCE Mesembryanthemum crystallinum
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
Caryophyllales; Caryophyllaceae; Aizoaceae; Mesembryanthemum.
1 (bases 1 to 576)
REFERENCE 1
AUTHORS Cushman, J.C.
TITLE An expressed sequence tag database for the common ice plant,
Mesembryanthemum crystallinum
JOURNAL Unpublished (1997)
COMMENT Contact: Cushman JC
Department of Biochemistry
University of Nevada
MS200, Reno, NV 89557-0014, USA
Tel: 775-784-1918
Fax: 775-784-1650
Email: jcushman@unr.edu
PCR Primers
FORWARD: T3
BACKWARD: T3
Plate: L48-14
Seq primer: T3
High quality sequence stop: 350
POLYA-No.

FEATURES
Source
1..576 Location/Qualifiers
/organism="Mesembryanthemum crystallinum"
/db_xref="taxon:3544"
/clone_lib="Ice plant Lambda Uni-Zap XR expression library
48 hours NacI treatment"
/tissue_type="leaf 48 h 0.4M NaCl"
/dev_stage="Six week old"
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
EcoRI; Site_2: XhoI"

```

```

BASE COUNT      152 a      141 c      118 g      165 t
ORIGIN

Query Match      70.4%; Score 17.6; DB 163; Length 576;
Best Local Similarity 83.3%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cattcgatcctgcagtgagc 25
||||||| 11111 11111
Db 465 CATTTCGACCATCTCTGCTCGAG 442

RESULT 12
BF682210 1032 bp mRNA EST 21-DEC-2000
LOCUS 602117637T1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
DEFINITION IMAGE:3468776 3', mRNA sequence.
ACCESSION BF682210
VERSION BF682210.1 GI:11956105
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1032)
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabs-r@mail.nih.gov
Tissue Procurement: M. Bento Soares, Ph.D.
CDNA Library Preparation: Bento Soares and M. Fatima Bonaldo
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
CDNA distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1A68476 row: j column: 09
High quality sequence start: 21
High quality sequence stop: 789.

FEATURES
Source
1..1032 Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3468776"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="Female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT73 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      268 a      261 c      218 g      265 t
ORIGIN

Query Match      70.4%; Score 17.6; DB 168; Length 1032;
Best Local Similarity 83.3%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cattcgatcctgcagtgagc 25
||||||| 11111 11111
Db 810 CATTTCATTCCTCTAGGTGAG 833

RESULT 13
BE595789/c

```

LOCUS BE595789 325 bp mRNA EST 18-AUG-2000
 DEFINITION P11.54.A07.b1.A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,
 mRNA sequence.
 ACCESSION BE595789
 VERSION BE595789.1 GI:9850862
 KEYWORDS EST.
 SOURCE Sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 325)
 Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
 L.H.
 REFERENCE An EST database from Sorghum: pathogen-induced plants
 AUTHORS Unpublished (2000)
 JOURNAL Contact: Cordonnier-Pratt MM
 COMMENT Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: emprat@uga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: JEN REV
 High quality sequence stop: 254
 POLYA-No.

FEATURES
 source
 1..325
 Location/Qualifiers
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Pathogen induced 1 (P11)"
 /note="Organ: Anthracnose-infected leaves from
 two-week-old sorghum plants 48 hr after inoculation;
 Vector: pBluescript II from Lambda Zap II; Site.1: XhoI;
 Site.2: EcoRI; Two-week-old sorghum plants (BRX 623
 cultivar) were infected with pathogen (isolate PM421 of
 Colletotrichum graminicola, which is a sorghum isolate).
 RNA was prepared from infected leaves harvested from 45
 seedlings 48 hours after inoculation. Note: young
 seedlings (2 weeks old) exhibit juvenile resistant
 reaction, which is an incompatible interaction. As they
 grow older (4 weeks or older), plants resume susceptibility
 to anthracnose disease. The library was made from poly-A
 RNA in the cloning vector lambda Zap II. Clones to be
 sequenced were prepared by mass excision. WARNING: While
 most or all ESTs are expected to derive from the host
 plant, no effort was made to eliminate ESTs deriving from
 the pathogen."

BASE COUNT 104 a 62 c 78 g 81 t
 ORIGIN

Query Match 68.8%; Score 17.2; DB 137; Length 325;
 Best Local Similarity 86.4%; Pred. No. 6e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 gcattgcacccgcgcagctg 22
 ||||| ||||| |||||
 Db 148 GCATTTGATCCTCATCAGAGAG 127

RESULT 14
 LOCUS AM604656 343 bp mRNA EST 23-MAR-2000
 DEFINITION CM0-CT0312-190100-153-d11 CT0312 Homo sapiens cDNA, mRNA sequence.
 ACCESSION AM604656
 VERSION AM604656.1 GI:7309397
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 343)
 AUTHORS HCGP <http://www.ludwig.org.br/ORESTES>.
 TITLE The FAPESP/LICR Human Cancer Genome Project
 JOURNAL Unpublished (1999)
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=CM0<2=CM0-CT0312-190100-153-d11<3=2000-01-19<4=1>)
 Seq primer: puc 18 forward
 High quality sequence start: 11
 High quality sequence stop: 116.

FEATURES
 source
 1..343
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="CT0312"
 /dev_stage="Adult"
 /note="Organ: colon; Vector: puc18; Site.1: SmaI; Site.2:
 SmaI; A mini-library was made by cloning products derived
 from ORESTES PCR (U.S. Letters Patent application No. 196
 716 - Ludwig Institute for Cancer Research) profiles
 into the pUC 18 vector. Reverse transcription of tissue
 mRNA and cDNA amplification were performed under low
 stringency conditions."

BASE COUNT 77 a 101 c 92 g 73 t
 ORIGIN

Query Match 68.8%; Score 17.2; DB 118; Length 343;
 Best Local Similarity 86.4%; Pred. No. 6.1e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 cattgcacccgcgcagctg 23
 ||||| ||||| |||||
 Db 252 CATTCGGTCATCTTCAGGTGC 231

RESULT 15
 LOCUS BF524069 349 bp mRNA EST 11-DEC-2000
 DEFINITION UI-R-C0-hs-f-07-0-UI.r1 UI-R-C0 Rattus norvegicus cDNA clone
 UI-R-C0-hs-f-07-0-UI 5', mRNA sequence.
 ACCESSION BF524069
 VERSION BF524069.1 GI:11632036
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 349)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu

CDNA Library Preparation: M.B. Soares Lab Clone distribution:
 clones will be available through Research Genetics (www.resgen.com)
 This clone is also available through the I.M.A.G.E. Consortium at
 LNL (info@image.lnl.gov). IMAGE ID= 1792330
 Seq primer: M13 forward.

FEATURES

source

Location/Qualifiers
 1. 349
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-C0-hs-f-07-00-UI"
 /clone_1lb="UI-R-C0"
 /dex_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C0
 library is a subtracted library derived from the UI-R-A1
 and UI-R-E1 libraries. The UI-R-A1 library consisted of a
 mixture of individually tagged normalized libraries
 constructed from rat placenta, adult lung, brain, liver,
 kidney, heart, spleen, ovary, and muscle. The UI-R-E1
 library consisted of a mixture of individually tagged
 normalized libraries constructed from 8, 12 and 18-day
 embryo. The tag is a string of 3-5 nucleotides present
 between the Not I site and the oligo-dT track which
 allows identification of the library of origin of a clone
 within the mixture. The subtracted library (UI-R-C0) was
 constructed as follows: PCR amplified cDNA inserts from a
 pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had
 been derived was used as a driver in a hybridization with
 the pooled UI-R-A1 and UI-R-E1 library in the form of
 single-stranded circles. The remaining single-stranded
 circles (subtracted library) was purified by
 hydroxyapatite column chromatography, converted to
 double-stranded circles and electroporated into DH10B
 bacteria (Life Technologies) to generate the UI-R-C0
 library. This procedure has been previously described
 (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
 1996)"

BASE COUNT 76 a 105 c 95 g 73 t
 ORIGIN

Query Match 68.8%; Score 17.2; DB 150; Length 349;
 Best Local Similarity 86.4%; Pred. No. 6.1e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 cattgatcctcgtcaagtgc 23
 |||||
 Db 295 CACTTCATCTCTCGTCAGGGGC 316

RESULT 16
 AF116319 364 bp DNA GSS 12-APR-2001
 LOCUS AF116319 Armillaria gallica W40f Armillaria gallica genomic, DNA
 DEFINITION sequence.
 ACCESSION AF116319
 VERSION AF116319.1 GI:4761841
 KEYWORDS GSS.
 SOURCE Armillaria gallica.
 ORGANISM Armillaria gallica
 Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Hymenomyetidae;
 Agaricales; Tricholomataceae; Armillaria.
 1 (bases 1 to 364)
 Hodnett, B. and Anderson, J. B.
 Genomic stability of two individuals of Armillaria gallica
 Mycologia 92 (5), 894-899 (2000)
 CONTACT: Anderson JB
 BIOLOGY
 University of Toronto at Mississauga
 3359 Mississauga Road, Mississauga, Ontario L5L 1C6, Canada
 Email: bhodnett@credit.erin.utoronto.ca

Class: unknown.

FEATURES

source

Location/Qualifiers
 1. 364
 /organism="Armillaria gallica"
 /strain="W40f"
 /db_xref="taxon:47427"
 /clone_1lb="Armillaria gallica W40f"
 BASE COUNT 56 a 107 c 71 g 128 t 2 others
 ORIGIN

Query Match 68.8%; Score 17.2; DB 218; Length 364;
 Best Local Similarity 86.4%; Pred. No. 6.1e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 cattgatcctcgtcaagtgc 23
 |||||
 Db 80 CATTTCCTTCAGTCAGGTC 101

RESULT 17
 AA946503/c 654 bp mRNA EST 08-JAN-1999
 LOCUS AA946503
 DEFINITION EST202002 Normalized rat ovary, Bento Soares Rattus sp. CDNA clone
 ROVAR89 3' end, mRNA sequence.
 ACCESSION AA946503.1 GI:4132826
 VERSION
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus
 1 (bases 1 to 654)
 Lee, N.H., Glodok, A., Chandra, I., Mason, T.M., Quackenbush, J.,
 Kerlavage, A.R. and Adams, M.D.
 Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat
 Gene Index
 Unpublished (1998)
 On May 1, 1998 this sequence version replaced gi:3106419.
 CONTACT: Lee, NH
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: M13-21.

FEATURES

source

Location/Qualifiers
 1. 654
 /organism="Rattus sp."
 /db_xref="taxon:10118"
 /clone="ROVAR89"
 /clone_1lb="Normalized rat ovary, Bento Soares"
 /note="Organ: ovary; Vector: pT73Pac; Site_1: EcoRI;
 Site_2: NotI"
 BASE COUNT 151 a 143 c 198 g 162 t
 ORIGIN

Query Match 68.8%; Score 17.2; DB 14; Length 654;
 Best Local Similarity 86.4%; Pred. No. 6.4e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 cattgatcctcgtcaagtgc 23
 |||||
 Db 575 CACTTCATCTCTCGTCAGGGGC 554

RESULT 18
 AW940604 719 bp mRNA EST 30-MAY-2000
 LOCUS AW940604
 DEFINITION GH15883.3 prline GH Drosophila melanogaster head pot2 Drosophila
 melanogaster cDNA clone GH15883 3, mRNA sequence.

ACCESSION AM940604
 VERSION GI:8116050
 KEYWORDS EST.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 719)
 REFERENCE Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein, P., Lewis, S. and Rubin, G.M.
 BDGP/HMI Drosophila EST Project
 Unpublished (1997)
 Other ESTs: GH15883, 5prime
 COMMENT Contact: Harvey, D.
 G. M. Rubin-Molecular and Cell Biology
 University of California Berkeley
 539 LSA, Berkeley, CA 94720-3200, USA
 Fax: 510 643 9947
 Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
 Based upon the presence of a XhoI site followed by a run of 14 or more T residues at the beginning of the sequence, this clone was polyadenylated. The resulting Poly-T sequence has been removed.
 Plate: GH.158 row: G column: 11
 High quality sequence stop: 718.
 FEATURES
 source Location/Qualifiers
 1..719
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="GH15883"
 /clone_lib="GH Drosophila melanogaster head pot2"
 /sex="male and female"
 /dev_stage="adult"
 /lab_host="DHS -alpha"
 /note="Organ: head; Vector: POT2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into POT2. Plasmid cDNA library."
 BASE COUNT 162 a 196 c 164 g 197 t
 ORIGIN

Query Match 68.8%; Score 17.2; DB 122; Length 719;
 Best Local Similarity 86.4%; Pred. No. 6.5e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 3 attcgatccctgcagtgca 24
 ||| |||| |||| |||| ||||
 Db 441 ATTGGATCTCGCGAGTGCA 462

RESULT 19
 CDS05JPO 988 bp DNA GSS 26-MAY-2000
 LOCUS Tetradon nigroviridis genome survey sequence T3 end of clone
 DEFINITION 028C23 of library A from Tetradon nigroviridis, genomic survey
 sequence.
 AL340487.1 GI:8234245
 VERSION GSS: genome survey sequence.
 KEYWORDS Tetradon nigroviridis.
 SOURCE Tetradon nigroviridis.
 ORGANISM Tetradon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetradon.
 1 (bases 1 to 988)
 REFERENCE Ruest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.
 Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis
 Unpublished
 2 (bases 1 to 988)
 JOURNAL
 TITLE
 AUTHORS
 REFERENCE
 COMMENT

AUTHORS Ruest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.
 TITLE Human gene number estimate provided by genome wide analysis using Tetradon nigroviridis DNA sequence
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 988)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetradon>.
 FEATURES
 source Location/Qualifiers
 1..988
 /organism="Tetradon nigroviridis"
 /db_xref="taxon:99883"
 /clone="028C23"
 /clone_lib="A"
 /note="Genoscope sequence ID : C0AA028BA12A2-end : T3"
 BASE COUNT 281 a 168 c 175 g 331 t 33 others
 ORIGIN

Query Match 68.8%; Score 17.2; DB 222; Length 988;
 Best Local Similarity 86.4%; Pred. No. 6.7e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 4 ttgcgctccctgcagtgag 25
 |||| |||| |||| ||||
 Db 796 TTTCGCTCTCTTCAGTGCAG 817

RESULT 20
 CDS05BPT 997 bp DNA GSS 26-MAY-2000
 LOCUS Tetradon nigroviridis genome survey sequence T7 end of clone
 DEFINITION 002M10 of library A from Tetradon nigroviridis, genomic survey
 sequence.
 AL330122.1 GI:8223744
 VERSION GSS: genome survey sequence.
 KEYWORDS Tetradon nigroviridis.
 SOURCE Tetradon nigroviridis.
 ORGANISM Tetradon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetradon.
 1 (bases 1 to 997)
 REFERENCE Ruest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.
 Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis
 Unpublished
 2 (bases 1 to 997)
 REFERENCE Ruest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.
 Human gene number estimate provided by genome wide analysis using Tetradon nigroviridis DNA sequence
 Unpublished
 3 (bases 1 to 997)
 JOURNAL
 TITLE
 AUTHORS
 REFERENCE
 COMMENT

Direct Submission
 Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetradon>.
 FEATURES
 source Location/Qualifiers
 1..997


```
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="002M10"
/note="Genoscope sequence ID : COAA002G05C1-end : T7"

BASE COUNT      260 a      222 c      193 g      296 t      26 others
ORIGIN

Query Match      68.8%; Score 17.2; DB 222; Length 997;
Best Local Similarity 86.4%; Pred. No. 6.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 4 ttgcgtcctcgtcagtgacg 25
    ||||| ||||| ||||| |||||
Db 71 TTTCGCTCCTCTTCAGTTCGAG 92

RESULT 21
LOCUS      CNS05670      1081 bp      DNA      GSS      26-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
043620 of library A from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION  ALJ350205.1 GI:8243975
VERSION     GSI: genome survey sequence.
KEYWORDS    Tetraodon nigroviridis.
SOURCE      Tetraodon nigroviridis
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE   1 (bases 1 to 1081)
AUTHORS    Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Weissenbach,J.,
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL     Unpublished
TITLE       2 (bases 1 to 1081)
AUTHORS     Roest-Crollius,H., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
COMMENT     Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
JOURNAL     3 (bases 1 to 1081)
REFERENCE   Genoscope.
AUTHORS     Direct Submission
TITLE       Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
JOURNAL     This sequence is a single read and was generated as part of a large
COMMENT     scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES
source
1..1081
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="043620"
/clone_1lb="A"
/note="Genoscope sequence ID : COAA043BD10C1-end : T7"

BASE COUNT      350 a      186 c      188 g      342 t      15 others
ORIGIN

Query Match      68.8%; Score 17.2; DB 222; Length 1081;
Best Local Similarity 86.4%; Pred. No. 6.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 4 ttgcgtcctcgtcagtgacg 25
    ||||| ||||| ||||| |||||
Db 776 TTTCGCTCCTCTTCAGTTCGAG 797

/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="002M10"
/note="Genoscope sequence ID : COAA002G05C1-end : T7"

BASE COUNT      260 a      222 c      193 g      296 t      26 others
ORIGIN

Query Match      68.8%; Score 17.2; DB 222; Length 997;
Best Local Similarity 86.4%; Pred. No. 6.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 4 ttgcgtcctcgtcagtgacg 25
    ||||| ||||| ||||| |||||
Db 71 TTTCGCTCCTCTTCAGTTCGAG 92

RESULT 22
LOCUS      CNS04VMF      1095 bp      DNA      GSS      24-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T3 end of clone
022N08 of library A from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION  ALJ309624.1 GI:8216563
VERSION     GSI: genome survey sequence.
KEYWORDS    Tetraodon nigroviridis.
SOURCE      Tetraodon nigroviridis
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE   1 (bases 1 to 1095)
AUTHORS    Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL     Unpublished
TITLE       2 (bases 1 to 1095)
AUTHORS     Roest-Crollius,H., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
COMMENT     Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
JOURNAL     3 (bases 1 to 1095)
REFERENCE   Genoscope.
AUTHORS     Direct Submission
TITLE       Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
JOURNAL     This sequence is a single read and was generated as part of a large
COMMENT     scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES
source
1..1095
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="022N08"
/clone_1lb="A"
/note="Genoscope sequence ID : COAA022G04A1-end : T3"

BASE COUNT      277 a      244 c      251 g      317 t      6 others
ORIGIN

Query Match      68.8%; Score 17.2; DB 221; Length 1095;
Best Local Similarity 86.4%; Pred. No. 6.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 4 ttgcgtcctcgtcagtgacg 25
    ||||| ||||| ||||| |||||
Db 596 TTTCAGTCTAGTCAGGAGGAG 617

RESULT 23
LOCUS      AA422533      113 bp      mRNA      EST      16-OCT-1997
DEFINITION vfi404.s1 Knowles Solter mouse unfertilized egg Mus musculus cDNA
clone IMAGE:835662.5' similar to SW:MNIN_SALTY P37169 VIRIDENCE
FACTOR MYIN. ;, mRNA sequence.
ACCESSION  AA422533.1 GI:2101353
VERSION     EST.
KEYWORDS    house mouse.
SOURCE      Mus musculus
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 113)
```

AUTHORS

Malina, W., Miller, L., Allen, M., Bowles, M., Dierlich, N., Dubuque, J., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE
JOURNAL
NUMBER

Contact: Maria M/Mouse EST Project
WashU-HIMM Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (infoimage.lnl.gov) for further information.
MGI:495878
Possible reversed clone: similarity on wrong strand
Seq primer: -40m3 fwd. RT from Amersham.

FEATURES

Source

BASE COUNT	15 a	32 c	36 g	30 t
ORIGIN				

Query Match	68.0%;	Score 17;	DB 7;	Length 113;
Best Local Similarity	80.0%;	Pred. No.	6.8e+02;	
Matches	20; Conservative	0;	Mismatches	5;
			Indels	0;
			Gaps	0
Dy	1 gcatctgcgacccgtcagtcagc	25		
Db	9 GCATGCCGACCTCTTCAGGTGG	33		

RESULT 24

LOCUS	BF386481	300 bp	EST	27-NOV-2000
DEFINITION	UI-R-CAL1-bbe-e-03-UI.s1 UI-R-CAL1 Rattus norvegicus CDNA clone			
ACCESSION	UI-R-CAL1-bbe-e-03-UI 3', mRNA sequence.			
VERSION	BF386481			
KEYWORDS	BF386481.1 GI:11371346			
SOURCE	EST			
ORGANISM	Norway rat.			
	Rattus norvegicus			

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

1 (bases 1 to 300)
 Ronaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 tel: 319 335 8250

FEATURES

Email: msquares@blue.weeg.uiowa.edu
Oligo-dT track not found. Not a site shown in beginning of sequence.
Is likely internal to the message. CDNA library preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 Forward
SOLYA=NO.

BASE COUNT
ORIGIN

56 a	81 c	77 g	86 t
------	------	------	------

Query	1	gcaatttcgacctctgcagcagcag	25
	11	11111111111111111111	
Db	27	gcctcttcttaccttctgcagctacag	51

Best local similarity 80.0%; Predicted No. 7.5e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

RESULT 25

LOCUS	A124043	332 bp	EST
DEFINITION	RH12L_E10_Y1_A001 Rhizome1		19-JUL-2000
DESCRIPTION	sequence.		Sorghum halpense cDNA, mRNA
ACCESSION	A124043		
VERSION	A1724043.1	GI:5042895	
KEYWORDS	EST.		
SOURCE	Sorghum halpense.		
ORGANISM	Sorghum halpense		

REFERENCE

AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Cordomier-Pratt, M.-M., Gingle, A., Pratt, L.H. and Paterson, A.
 An EST database from Sorghum: Sorghum halepense rhizomes
 Unpublished (2000)
 Contact: Cordomier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel.: 706 542 1860

Sequences have been trimmed to exclude POLYA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: T7
High quality sequence start: 4
High quality sequence stop: 332
POLYA=Yes.

FEATURES

Source

```

/organism="Sorghum halepense"
/db_xref="taxon:4560"
/clone_lib="Rhizomes (RHIZ1)"
/note="Organ: Rhizomes; Vector: Bluescript II from Lambda
Zap II; Site: 1; XhoI; Site: 2; EcoRI; The library was made
from poly-A RNA in the cloning vector Lambda Zap II.
Clones to be sequenced were prepared by mass excision."
BASE COUNT      74 a      93 c      75 g      90 t
ORIGIN

Query Match      68.0%; Score 17; DB 24; Length 332;
Best Local Similarity 80.0%; Pred. No. 7.5e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gcattcgatccctgcagtgacg 25
|||||
Db 131 GCATGTCGATCCTCTCTGAGCGG 107

RESULT 26
A2817869      360 bp      DNA      GSS      20-FEB-2001
LOCUS      2M0087N04R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION      clone UUGC2M0087N04 R. DNA sequence.
ACCESSION      A2817869
VERSION      A2817869.1 GI:12987777
KEYWORDS      GSS.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murineae; Mus.
REFERENCE      1 (bases 1 to 360)
AUTHORS      Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE      Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL      Unpublished (2000)
COMMENT      Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0087 row: N column: 04
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 360.
Location/Qualifiers
1. 360
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0087N04"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel

```

```

electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (g1473211419b1AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      102 a      79 c      95 g      83 t      1 others
ORIGIN

Query Match      68.0%; Score 17; DB 250; Length 360;
Best Local Similarity 80.0%; Pred. No. 7.6e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gcattcgatccctgcagtgacg 25
|||||
Db 102 GCATTCGACGCTCTGAGCTCAG 126

RESULT 27
AW747232      421 bp      mRNA      EST      19-JUL-2000
LOCUS      AW747232/c
DEFINITION      W51.66.E04.g1.A002 Water-stressed 1 (W51) Sorghum bicolor cDNA,
mRNA sequence.
ACCESSION      AW747232
VERSION      AW747232.1 GI:7660970
KEYWORDS      EST.
SOURCE      sorghum.
ORGANISM      Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE      1 (bases 1 to 421)
AUTHORS      Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
,L.H.
TITLE      An EST database from Sorghum: water-stressed plants
JOURNAL      Unpublished (2000)
COMMENT      Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: T7
High quality sequence stop: 409
PolyA=yes.
Location/Qualifiers
1. 421
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Water-stressed 1 (W51)"
/note="Organ: Mix of 5-week old plants on days 7 & 8 after
water was withheld; Vector: Lambda Zap; Site: 1; XhoI;
Site: 2; EcoRI; The library was made from poly-A RNA in the
cloning vector Lambda Zap II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT      96 a      120 c      111 g      94 t
ORIGIN

Query Match      68.0%; Score 17; DB 120; Length 421;
Best Local Similarity 80.0%; Pred. No. 7.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gcattcgatccctgcagtgacg 25
|||||
Db 244 GCATGTCGATCCTCTGAGCGG 220

```


found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bdrrp/image/image.html

Trace considered overall poor quality
Insert Length: 2267 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 1.

FEATURES

SOURCE

1. 456

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_image="1865314"

/clone_lib="NCI-CGAP_Kid3"

/lab_host="DH10B"

/note="Organ: Kidney; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer, (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. mRNA source: 2 pooled kidneys. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

124 a 128 c 110 g 94 t

ORIGIN

Query Match 68.0%; Score 17; DB 18; Length 456;

Best Local Similarity 80.0%; Pred. No. 7.7e+02;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gcaattcgatcctcgtagtgag 25

||||| ||||| ||||| |||||

Db 260 GCATCTGCATCTTCATCTGCAG 236

RESULT 31

LOCUS

BG240831 463 bp mRNA EST 15-FEB-2001

DEFINITION BG240831.H01.g1_A002 Ovary 1 (OVI) Sorghum bicolor cDNA, mRNA

ACCESSION

BG240831 GI:12775904

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

An EST database from Sorghum: ovaries of varying immature stages

Unpublished (2000)

Contact: Cordomier-Pratt MM

Department of Botany

The University of Georgia

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 542 1805

Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions

below Phred quality 16. The threshold for highest quality sequence

is 20.

Seq primer: PolYTmIX

High quality sequence start: 2

High quality sequence stop: 461

POLYA-No

Location/Qualifiers

1. 463

/organism="Sorghum bicolor"

/db_xref="taxon:4558"

/clone_lib="Ovary 1 (OVI)"

FEATURES

SOURCE

/note="Organ: Mix of ovaries of varying immature stages from 8-week-old plants; Vector: pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

107 a 131 c 124 g 100 t 1 others

BASE COUNT

107 a 131 c 124 g 100 t 1 others

ORIGIN

Query Match 68.0%; Score 17; DB 175; Length 463;

Best Local Similarity 80.0%; Pred. No. 7.8e+02;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gcaattcgatcctcgtagtgag 25

Db 300 GCATGTGATCTCTCTGAGCCG 276

RESULT 32

LOCUS

BG047654 471 bp mRNA EST 25-JAN-2001

DEFINITION BG047654.E07.b1_A002 Ovary 1 (OVI) Sorghum bicolor cDNA, mRNA

ACCESSION

BG047654 GI:12497673

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

An EST database from Sorghum: ovaries of varying immature stages

Unpublished (2000)

Contact: Cordomier-Pratt MM

Department of Botany

The University of Georgia

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 542 1805

Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions

below Phred quality 16. The threshold for highest quality sequence

is 20.

Seq primer: JEN REV

High quality sequence stop: 467

POLYA-No.

Location/Qualifiers

1. 471

/organism="Sorghum bicolor"

/db_xref="taxon:4558"

/clone_lib="Ovary 1 (OVI)"

/note="Organ: Mix of ovaries of varying immature stages from 8-week-old plants; Vector: pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

BG048007/c
 LOCUS 473 bp mRNA EST 25-JAN-2001
 DEFINITION OV1.26.E07.gL.A002 Ovary 1 (OV1) Sorghum bicolor cDNA, mRNA
 ACCESSION BG048007
 VERSION BG048007.1 GI:12498325
 KEYWORDS EST.
 SOURCE sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 473)
 Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
 ,L.H.
 An EST database from Sorghum: ovaries of varying immature stages
 Unpublished (2000)
 CONTACT: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: PolYTmIX
 High quality sequence stop: 465
 POLYA-No.

FEATURES
 source Location/Qualifiers
 1..473
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Ovary 1 (OV1)"
 /note="Organ: Mix of ovaries of varying immature stages
 from 8-week-old plants; Vector: pLuscript II from Lambda
 Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
 from polyA RNA in the cloning vector Lambda Zap II.
 Clones to be sequenced were prepared by mass excision."
 BASE COUNT 109 a 134 c 127 g 103 t
 ORIGIN

Query Match 68.0%; Score 17; DB 173; Length 473;
 Best Local Similarity 80.0%; Pred. No. 7.8e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gcaattcgatcctgcagtcgacg 25
 ||| ||||| ||| ||| |||
 Db 300 GCATGTCATCTCCTCTGACGCG 276

RESULT 34
 BF388049
 LOCUS 476 bp mRNA EST 27-NOV-2000
 DEFINITION UT-R-CA1-bbx-f-10-0-UI-sl UT-R-CA1 Rattus norvegicus cDNA clone
 ACCESSION BF388049
 VERSION BF388049.1 GI:11372873
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 476)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 JOURNAL MEDLINE 97044477
 COMMENT Contact: Soares, MB

Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: mscoates@blue.weeg.uiowa.edu
 Oligo-dT track not found, Not 1 site shown in beginning of sequence
 is likely internal to the message. cDNA Library Preparation: M.B.
 Soares Lab clone distribution: clones will be available through
 Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA-No.

FEATURES
 source Location/Qualifiers
 1..476
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone_lib="O1-R-CA1"
 /clone_1b="O1-R-CA1"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: p773D-Pac (Pharmacia)"
 polylinker: Site_1: Not I; Site_2: Eco RI; The UT-R-CA1
 library is a subtracted library derived from the following
 tissues: thalamus, cerebellum, hypothalamus, medulla, pons
 and hippocampus. For a detailed description of the library
 and hippocampus. For a detailed description of the library
 from which this clone was derived, please visit our web
 site at rat.est.eng.uiowa.edu. The subtraction has been
 previously described in (Bonaldo, Lennon and Soares,
 Genome Research 6:791-806, 1996)
 TAG-SRO=None found"

BASE COUNT 81 a 117 c 125 g 150 t 3 others
 ORIGIN

Query Match 68.0%; Score 17; DB 148; Length 476;
 Best Local Similarity 80.0%; Pred. No. 7.8e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gcaattcgatcctgcagtcgacg 25
 ||| ||||| ||||| ||| |||
 Db 27 GCCTCTCTATCTGTCGACGTACAG 51

RESULT 35
 AM923923
 LOCUS 489 bp mRNA EST 19-JUL-2000
 DEFINITION WSI_30.A09.b1.A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
 mRNA sequence.
 ACCESSION AM923923
 VERSION AM923923.1 GI:8089748
 KEYWORDS EST.
 SOURCE Sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 489)
 Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
 ,L.H.
 An EST database from Sorghum: water-stressed plants
 Unpublished (2000)
 CONTACT: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: JEN REV

High quality sequence stop: 472
POLYA-No.

FEATURES
SOURCE
Location/Qualifiers
1. .489

/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Water-stressed 1 (WS1)"
/note="Organ: Mix of 5-week old plants on days 7 & 8 after water was withheld; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."
BASE COUNT
120 a 136 c 125 g 108 t
ORIGIN

Query Match 68.0%; Score 17; DB 122; Length 489;
Best Local Similarity 80.0%; Pred. No. 7.8e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 gcatctgacatcctcctcctgagccg 25
|||||
Db 343 GCATGTCGATCCTCCTCTGAGCCG 319

RESULT 36
BE355471 497 bp mRNA EST 20-JUL-2000
LOCUS DGL114.H10.g1_A002 Dark Grown 1 (DGL) sorghum bicolor cDNA, mRNA
DEFINITION
sequence.
ACCESSION BE355471
VERSION BE355471.1 GI:9296573
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Sorghum.
REFERENCE
1 (bases 1 to 497)
Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and Pratt
, L.H.

TITLE An EST database from Sorghum: dark-grown seedlings
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude polyA, vector and regions
below phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: PolyTMMix
High quality sequence start: 2
High quality sequence stop: 475
POLYA-No.
Location/Qualifiers
1. .497

/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DGL)"
/note="Organ: 5-day-old dark-grown seedlings; Vector:
Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
made from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced were prepared by mass excision."
BASE COUNT
125 a 137 c 123 g 111 t
ORIGIN

Query Match 68.0%; Score 17; DB 166; Length 497;
Best Local Similarity 80.0%; Pred. No. 7.8e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gcatctgacatcctcctcctgagccg 25
|||||
Db 335 GCATGTCGATCCTCCTCTGAGCCG 311

RESULT 37
AW747172 502 bp mRNA EST 19-JUL-2000
LOCUS AW747172/c
DEFINITION
WS1_66_E04.b1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
mRNA sequence.
ACCESSION AW747172
VERSION AW747172.1 GI:7660910
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Sorghum.
REFERENCE
1 (bases 1 to 502)
Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and Pratt
, L.H.
An EST database from Sorghum: water-stressed plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude polyA, vector and regions
below phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 458
POLYA-No.
Location/Qualifiers
1. .502

/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Water-stressed 1 (WS1)"
/note="Organ: Mix of 5-week old plants on days 7 & 8 after water was withheld; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."
BASE COUNT
132 a 136 c 130 g 104 t
ORIGIN

Query Match 68.0%; Score 17; DB 120; Length 502;
Best Local Similarity 80.0%; Pred. No. 7.8e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gcatctgacatcctcctcctgagccg 25
|||||
Db 304 GCATGTCGATCCTCCTCTGAGCCG 280

RESULT 38
A2475043 506 bp DNA GSS 04-OCT-2000
LOCUS A2475043
DEFINITION
JM0292H17R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0292H17 R, DNA sequence.
ACCESSION A2475043
VERSION A2475043.1 GI:10633168
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 506)

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmp@tuga.edu

Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: JEN REV

High quality sequence stop: 503

POLYA-No.

FEATURES

source

Location/Qualifiers

1..511

/organism="Sorghum bicolor"

/db_xref="taxon:4558"

/clone_lib="Pathogen induced 1 (PI1)"

/note="Organ: Anthracnose-infected leaves from two-week-old sorghum plants 48 hr after inoculation; Vector: pBluescript II from lambda Zap II; Site 1: XhoI; Site 2: EcoRI; Two-week-old sorghum plants (BRX 623 cultivar) were infected with pathogen (isolate FRM421 of Colletotrichum graminicola, which is a sorghum isolate). RNA was prepared from infected leaves harvested from 45 seedlings (2 weeks old) exhibit juvenile resistant reaction, which is an incompatible interaction. As they grow older (4 weeks or older), plants resume susceptibility to anthracnose disease. The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."

BASE COUNT

108 a 138 c 116 g 147 t 2 others

ORIGIN

Query Match 68.0%; Score 17; DB 166; Length 511;
Best Local Similarity 80.0%; Pred. No. 7.8e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query 1 gcaattgcattcctcgtcagtgca 25
|||||
Db 191 GCATGTCGATCTCTCTCGAGCCG 167

RESULT 41

BE364141/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

Clade: Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 511)

CDONNIE-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt

,L.H.

An EST database from Sorghum: pathogen-induced plants

Unpublished (2000)

Contact: Cordonnier-Pratt MM

Department of Botany

The University of Georgia

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 542 1805

Email: mmp@tuga.edu

Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: JEN REV

High quality sequence stop: 503

POLYA-No.

Location/Qualifiers

1..511

/organism="Sorghum bicolor"

Seq primer: T7
High quality sequence stop: 511
POLYA-yes.

FEATURES

source

Location/Qualifiers

1..511

/organism="Sorghum bicolor"

/db_xref="taxon:4558"

/clone_lib="Pathogen induced 1 (PI1)"

/note="Organ: Anthracnose-infected leaves from two-week-old sorghum plants 48 hr after inoculation; Vector: pBluescript II from lambda Zap II; Site 1: XhoI; Site 2: EcoRI; Two-week-old sorghum plants (BRX 623 cultivar) were infected with pathogen (isolate FRM421 of Colletotrichum graminicola, which is a sorghum isolate). RNA was prepared from infected leaves harvested from 45 seedlings (2 weeks old) exhibit juvenile resistant reaction, which is an incompatible interaction. As they grow older (4 weeks or older), plants resume susceptibility to anthracnose disease. The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."

BASE COUNT

108 a 138 c 116 g 149 t

ORIGIN

Query Match 68.0%; Score 17; DB 166; Length 511;
Best Local Similarity 80.0%; Pred. No. 7.8e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query 1 gcaattgcattcctcgtcagtgca 25
|||||
Db 191 GCATGTCGATCTCTCTCGAGCCG 167

RESULT 42

A0940100

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Trypanosoma brucei

Eukaryota; Eukaryozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

1 (bases 1 to 515)

El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,

Gerard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,

Fraser,C. and Adams,M.

Determination of clone end sequences from Trypanosoma brucei GUTat

10.1 sheared DNA library

Unpublished (1999)

Other-GSSs: Sheared DNA-43C24.TF

Contact: Najib M. El-Sayed

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: nelsayed@tigr.org

Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared

DNA library constructed at TIGR. Clones will be available for

distribution through ATCC. Sheared DNA end sequences search page:

http://www.tigr.org/cdb/mdb/cdbdb/.

Seq primer: M13-Reverse

Class: Shotgun.

Location/Qualifiers

1..515

/organism="Trypanosoma brucei"

/db_xref="taxon:5828"

/clone_lib="GUTat 10.1 sheared DNA library"

```

/organism="Trypanosoma brucei"
/db_xref="taxon:5691"
/clone_lib="Sheared DNA-43C24"
/Note="Vector: pUC18; Site_1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TR927/4 Gnat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. & Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrel, Oxford University Press, 1999)."

```

BASE COUNT	106 a	158 c	119 g	132 t
ORIGIN				
Query Match				
Best Local Similarity	68.0%;	Score 17;	DB 235;	Length 515;
Matches 20;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0
Oy	1	gcatttcaccccgtagtgagcag 25		
Db	99	GCCTTTGATCCGCCCTCCGTGACG 123		
RESULT 43				
BE355404/c	518 bp	mRNA	EST	20-JUL-2000
LOCUS	DGI.114.H10.b2_A002	Dark Grown 1 (DGI)	Sorghum bicolor cDNA, mRNA	
DEFINITION	Sequence.			
ACCESSION	BE355404			
VERSION	BE355404.1	GI:9296506		
KEYWORDS	EST.			
SOURCE	sorghum.			
ORGANISM	Sorghum bicolor			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.			
REFERENCE	1 (bases 1 to 518)			
AUTHORS	Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt ,L.H.			
TITLE	An EST database from Sorghum: dark-grown seedlings			
JOURNAL	Unpublished (2000)			
COMMENT	Contact: Cordonnier-Pratt MM Department of Botany The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel.: 706 542 1860 Fax: 706 542 1805 Email: mmp@prattuga.edu			
	Sequences have been trimmed to exclude polyA, vector and regions below phred quality 16. The threshold for highest quality sequence is 20. Seq primer: JEN REV High quality sequence stop: 511			
POLYA-NO.				
FEATURES				
SOURCE				
	1..518			
	/organism="Sorghum bicolor"			
	/db_xref="taxon:4558"			
	/clone_lib="Dark Grown 1 (DGI)"			
	/note="Organ: 5-day-old dark-grown seedlings; Vector: lambda zap; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."			
BASE COUNT	130 a	149 c	126 g	113 t
ORIGIN				

Query Match	68.0%	Score 17	DB 166	Length 518
Best Local Similarity	80.0%	Pred. No. 7.8e+02		
Matches 20	Conservative 0	Mismatches 5	Indels 0	Gaps 0
QY	1	gcatttcgattccctgcgtcagtgacg	25	
Db	351	GCATGTCGATCTCTCTGTGAGCGC	327	
RESULT 44				
BE35326/c	519 bp	mrna	EST	20-JUL-2000
LOCUS	DG1_114_b1_A002	Dark Grown 1 (DG1)	Sorghum bicolor cDNA, mRNA	
DEFINITION	sequence.			
ACCESSION	BE35326			
VERSION	BE35326.1	GI:9296428		
KEYWORDS	EST.			
SOURCE	sorghum.			
ORGANISM	Sorghum bicolor			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC			
	clade; Panicoideae; Andropogoneae; Sorghum.			
REFERENCE	1 (bases 1 to 519)			
AUTHORS	Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt			
	,L.H.			
TITLE	An EST database from Sorghum: dark-grown seedlings			
JOURNAL	Unpublished (2000)			
COMMENT	Contact: Cordonnier-Pratt MM			
	Department of Botany			
	The University of Georgia			
	Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA			
	Tel: 706 542 1860			
	Fax: 706 542 1805			
	Email: mmp@atl@uga.edu			
	Sequences have been trimmed to exclude PolyA, vector and regions			
	below phred quality 16. The threshold for highest quality sequence			
	is 20.			
	Seq primer: JEN REV			
	High quality sequence stop: 512			
	POLYA=NO.			
FEATURES				
SOURCE	Location/Qualifiers			
	1..519			
	/organism="Sorghum bicolor"			
	/db_xref="taxon:4558"			
	/clone_lib="Dark Grown 1 (DG1)"			
	/note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."			
BASE COUNT	130 a 149 c 127 g 113 t			
ORIGIN				
	Query Match	68.0%	Score 17	DB 166
	Best Local Similarity	80.0%	Pred. No. 7.8e+02	
	Matches 20	Conservative 0	Mismatches 5	Indels 0
OY	1	gcatttcgattccctgcgtcagtgacg	25	
Db	351	GCATGTCGATCTCTCTGTGAGCGC	327	
RESULT 45				
TA367G09Q	529 bp	DNA	GSS	13-DEC-2000
DEFINITION	T. brucei sheared genomic DNA clone 367g09, reverse sequence,			
ACCESSION	AL495254			
VERSION	AL495254.1	GI:11871641		
KEYWORDS	GSS.			
SOURCE	Trypanosoma brucei.			
ORGANISM	Trypanosoma brucei			
	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;			

REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT

Trypanosoma.
1 (bases 1 to 529)
Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk
Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at <http://www.sanger.ac.uk/Projects/T-brucei/>.

FEATURES

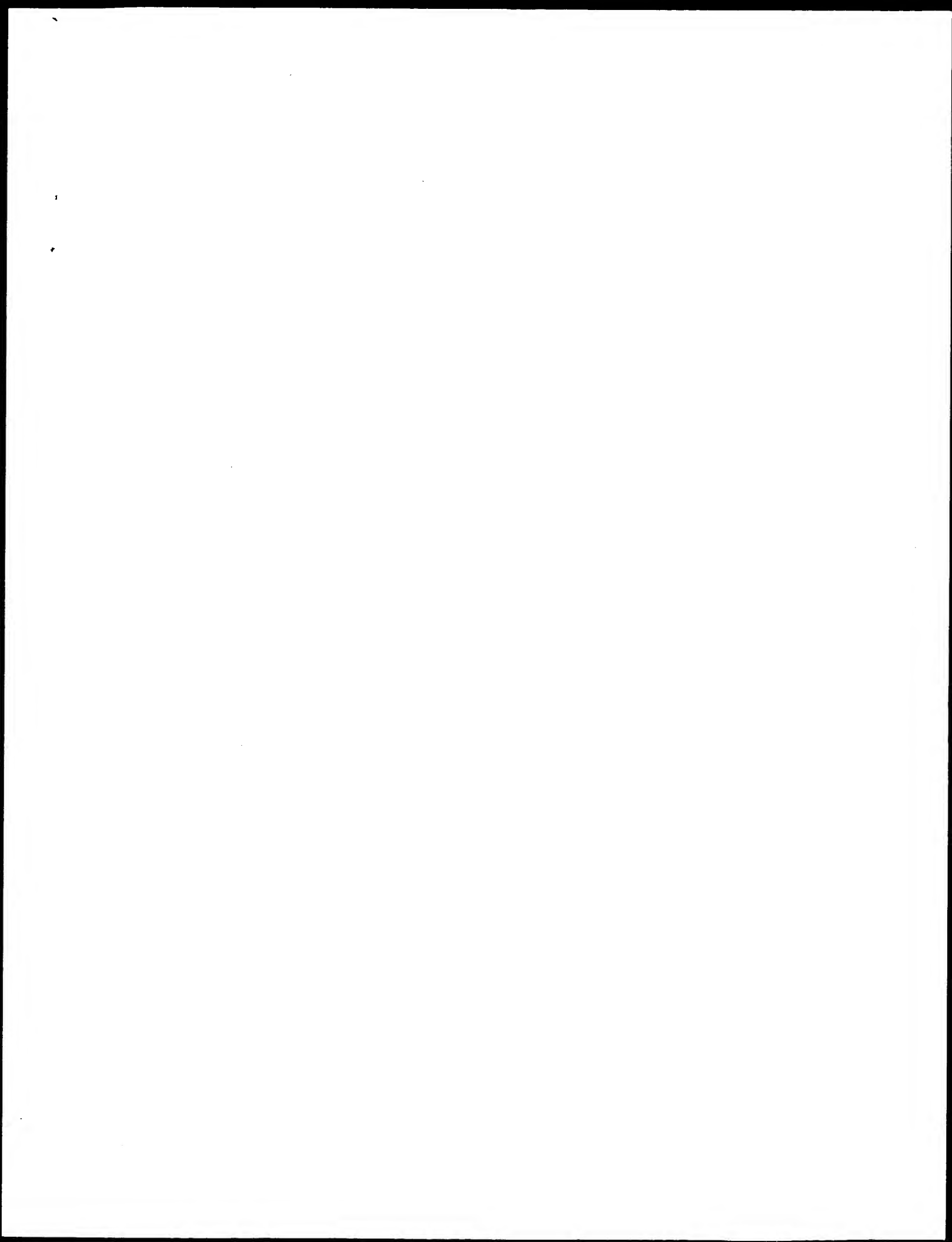
source

1..529
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="367909"

BASE COUNT 117 a 167 c 114 g 131 t
ORIGIN

Query Match 68.0%; Score 17; DB 258; Length 529;
Best Local Similarity 80.0%; Pred. No. 7.8e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 gcatcgcacccgcgcagtgacg 25
|| ||||| ||||| ||||| |||||
DB 455 GCCTTGATCCGCTCCGCTGCAG 479

Search completed: October 9, 2001, 13:46:49
Job time: 9524 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 12:14:30 ; Search time 1666.31 Seconds
(without alignments)
232.066 Million cell updates/sec

Title: US-09-396-196f-9
Perfect score: 25
Sequence: 1 cgatcctcgtcaggtcaggtcagc 25

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1344157 seqs, 773874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da1:*
2: gb_da2:*
3: gb_da3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: gb_da1:*
17: gb_da2:*
18: gb_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_r01:*
95: gb_r02:*
96: gb_r04:*
97: gb_pr10:*
98: em_da3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	9 AR029499	AR029499 Sequence
2	25	100.0	1041	9 AR034916	AR034916 Sequence
3	25	100.0	1084	9 A11530	A11530 BioB gene O
4	25	100.0	1121	10 E00893	E00893 Genomic DNA
5	25	100.0	5793	2 E00810	J04423 E.coli 7,8-
6	25	100.0	5872	9 A38246	A38246 Sequence 1
7	25	100.0	5872	9 A38251	A38251 Sequence 6
8	25	100.0	5872	9 A93674	A93674 Sequence 1

9 25 100.0 5872 9 A93679
10 25 100.0 5872 9 ARI01809
11 25 100.0 5872 9 ARI01810
12 25 100.0 11022 9 AE000180
13 25 100.0 13501 1 AE005253
14 25 100.0 297816 2 AE005253
15 23.4 93.6 5526 3 AF250776
16 21.8 87.2 7215 3 AF250776
17 20.8 83.2 4000 2 AF250768
18 20.2 80.8 1041 9 HSY15922
19 18.2 72.8 160 97 HSY15922
20 18.2 72.8 987 9 CEU10160
21 18.2 72.8 987 9 AR008950
22 18.2 72.8 987 10 128591
23 18.2 72.8 993 7 CEU57752
24 18.2 72.8 1005 91 CEU19834
25 18.2 72.8 1007 89 AF180563
26 18.2 72.8 1012 7 AF007576
27 18.2 72.8 1012 7 BTU11815
28 18.2 72.8 1018 9 AR052857
29 18.2 72.8 1058 7 ECIL12P40
30 18.2 72.8 1080 97 MMU19841
31 18.2 72.8 1389 10 AX084106
32 18.2 72.8 1399 9 AR091391
33 18.2 72.8 1399 97 HOMCLMF40
34 18.2 72.8 1560 9 AR091395
35 18.2 72.8 1623 9 AR091394
36 18.2 72.8 1870 9 A92079
37 18.2 72.8 2318 97 HOMKNSFP40
38 18.2 72.8 2362 9 AR082684
39 18.2 72.8 2362 10 128325
40 18.2 72.8 2362 10 189770
41 18.2 72.8 2611 93 HSAJ2967
42 18.2 72.8 6139 9 AR091393
43 18.2 72.8 18340 1 AY008847
44 18.2 72.8 33692 61 AC010294
45 18.2 72.8 42729 3 MICY71

ALIGNMENTS

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D,Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
source location/Qualifiers
1..1041
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatcctcgatcaggtgcagtcacg 25
|||||
Db 105 CGATCCTCGTCAGGTGCAGTCACG 129

RESULT 2

AR034916 AR034916 1041 bp DNA PAT 29-SEP-1999
LOCUS AR034916
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
source location/Qualifiers
1..1041
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatcctcgatcaggtgcagtcacg 25
|||||
Db 105 CGATCCTCGTCAGGTGCAGTCACG 129

RESULT 3
LOCUS A11530 1084 bp DNA PAT 09-FEB-1994
DEFINITION BioB gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 1084)
AUTHORS Bacteri; Proteobacteria; gamma subdivision; Enterobacteriaceae;
JOURNAL Escherichia.
FEATURES
source location/Qualifiers
1..1084
/organism="Escherichia coli"
/db_xref="taxon:562"
/gene="BIOB"
/cdon_start=1
/transl_table=1
/protein_id="CAA00965.1"
/db_xref="GI:490219"
/translation="MAHRPRMTLSQVTLFEKPLLDLFEAQYVHRQHFDPVOVST
LISITGACPECKTCPOSSRYKTELEERIMEVEOYLESRKAAGSTRFCMGAM
KNHEDMPVLEQWQVAKAGLEACMTLGLSSOARLNAGLDYVNHNLDSPP
YGNITTRTYOERDLTEKVDAGIKVCSGIVIGETVRAGLIDYLANLP
VPINMLVVKSTPLADNDVDAPFIRIYAVRIMMPSYVRLSAGREONNEOQAMC
FMAGNSIFCYGCKLITTPNPEEDKDLQLEFRKLINPOOTAVLAGDNEOQRLBOALMT
PDLEYYVAAL"

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatcctcgatcaggtgcagtcacg 25
|||||


```

/codon_start=1
/transl_table=11
/protein_id="AAA23517.1"
/db_xref="GI:145427"
/translation="MATVVKQAIYAARGRAAAHYEQHADLQROSDALLAMPORRYT
HYLDAGCGPGRMRRERHQAQVATLADSPMLVOAQKDAADHYLAGDIESLPLATA
FEDLAMSNIAYQMGCSNLTALRELYRVAPRGVAAFTLVQSGIPERHQMQAVDERP
HANRFLPDEIEOSLNGVHYOHHTOPTIMPDALSMRSLKIGATHLHGDRPIL
TRSOLORIOLAMPQOQGRYPLTYHLFGVIARE"
4938..5597
/gene="biob"
/locus_tag="4938..5597"
/gene="biob"
/codon_start=1
/transl_table=11
/product="gethioflavin synthetase"
/protein_id="AAA23518.1"
/db_xref="GI:145428"
/translation="MSKRYFVTGDTVEGKTVASCALLQAKRAGYRTAGYKPVASGS
EKTEGCRNSDALIORNSLQDVAIVNYPTEAPSPILVIAOGRPIESLVMASG
LRALQOADMVLYGAGGWFPLSDPTEFADWTOLOPILLYVGYKLCIINHMLTA
QVIOHACGLTLAGVANDVTPPGKRHAHEYMTTLTRMLPRCWEKSPGLQIKMQRPES
"
BASE COUNT      1363 a 1554 c 1631 g 1245 t
ORIGIN      4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match      100.0%; Score 25; DB 2; Length 5793;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1      cgatccctgcaggtgcagtcagc 25
Db      2116      CGATCCTGCTCAGGTGCGAGGTCAGC 2140

RESULT      6
LOCUS      A38246      5872 bp      DNA
DEFINITION      Sequence 1 from Patent WO9408023.
ACCESSION      A38246
VERSION      A38246.1      GI:2294844
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
      Escherichia.
      Escherichia coli.
      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
      Escherichia.
      1 (bases 1 to 5872)
      Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
      BIOLOGICAL METHOD OF PRODUCING BIOTIN
      Patent: WO 9408023-A 1 14-APR-1994;
      LONZA AG (CH)
COMMENT      Other publication PL 308301 950724
      Other publication CA 2145400 940414
      Other publication AU 4820293 940426
      Other publication HU 71781 960228
      Other publication SK 42095 951108
      Other publication CZ 9500809 950913
      Other publication FI 951547 950331
      Other publication JP 85016947 960227.
      Location/Qualifiers
        1..5872
          /organism="Escherichia coli"
          /strain="DSM498"
          /db_xref="taxon:562"
          /clone="PBO30A-15/9"
          1..96
            /function="PROMOTER PTAC"
            /evidence=experimental
            23..28
              /evidence="experimental"
              45..49
                /standard_name="PROMOTER PTAC"
                /evidence=experimental

FEATURES
source
promoter
-35_signal
-10_signal

```

```

RBS      105..119
      /standard_name="BIOB RBS NO.9"
      /evidence=experimental
gene      117..1157
      /gene="BIOB"
      117..1157
      /gene="BIOB"
      /codon_start=1
      /transl_table=11
      /number=1
      /evidence=experimental
      /product="BIOTIN SYNTHASE"
      /protein_id="CAA02325.1"
      /db_xref="GI:2294845"
      /translation="MAHRPMTLSQVTELEFKRILDLFEAQOVRHOFDPQOVST
      LLSITGACPEDCKYCRQSSRYRTGIEAERLSEVLESARKARAAGSTRCMAAM
      KNHERPMRYLEQWQGVKAMGLEACMTVSTLSESOARLANAGADLYNNHNLDSPEF
      YGNTITRYOERLDLLEKVRAGIVGSGGIVIGETVKDRAGLLLOLANLPPEPES
      VPINMLKVGSTPLADNDVDAEDFIRTAIVARIMPTSYVRLSAGROMNPOQAMC
      PMAGANSIEFGCKLTTPNPEEDKDLQFRKLGILNPQOTAVLAGDNEQORLEQALMT
      PDIDEXYNAAL"
      2284..2297
      /standard_name="BIOC RBS"
      2295..3050
      /gene="BIOC"
      2295..3050
      /gene="BIOC"
      /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
      /codon_start=1
      /transl_table=11
      /number=3
      /product="PROTEIN"
      /protein_id="CAA02326.1"
      /db_xref="GI:2294846"
      /translation="MATVVKQAIYAARGRAAAHYEQHADLQROSDALLAMPORRYT
      HYLDAGCGPGRMRRERHQAQVATLADSPMLVOAQKDAADHYLAGDIESLPLATA
      FEDLAMSNIAYQMGCSNLTALRELYRVAPRGVAAFTLVQSGIPERHQMQAVDERP
      HANRFLPDEIEOSLNGVHYOHHTOPTIMPDALSMRSLKIGATHLHGDRPIL
      TRSOLORIOLAMPQOQGRYPLTYHLFGVIARE"
      3742..3752
      /standard_name="BIOA RBS"
      3750..5039
      /gene="BIOA"
      3750..5039
      /gene="BIOA"
      /gene="BIOA"
      /standard_name="S-ADENOSYL-L-METHIONINE:
      8-AMINO-7-OXONONANOATE AMINOTRANSF."
      /EC_number="2.6.1.62"
      /codon_start=1
      /transl_table=11
      /number=5
      /evidence=experimental
      /product="DAPA SYNTHASE"
      /protein_id="CAA02327.1"
      /db_xref="GI:2294847"
      /translation="MTTDLAFDORHIMHTSMTSPYIPYVSAEGCELLISDGR
      LEQVFLADSGSVAEYVAKMAIQTWQKGFARQPLTFRNGYHGFAGMSVCDPDS
      MHSIMKGTLPENLPAAPQSRMDGEMDERUMVGFARLMAHREHIAAVTIEIYVAG
      GMRWHPBWLKRIKICDREGILLADELATGFRGKLFACSHAEIADITLCKAL
      QVADIEVLEQIOLAPADAEVADAVRVLGIGVETTHVYMAALQKFFVEGCVIRP
      FGKLIYMPYITLPPQQLRLTAAVRAVODEFFQO"
      5088..5100
      /standard_name="ORFI RBS"
      5098..5574
      /gene="ORFI"
      5098..5574
      /gene="ORFI"
      /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
      /codon_start=1
      /transl_table=11
      /number=6

```



```
/evidence=experimental
/product="PROTEIN"
/protein_id="CAA02328.1"
/db_xref="GI:2294848"
/translation="MKLISNDLRGDKLPHRHVNGMGYDQDNISPLHNDYPAKIG
SEFVVCYDPAATGSGMHWVYVNLPAATRYLPDGFSGGLVAMPDGLQRTDFKIG
YDGAAPKGGETHRYFTVHALDIERIDVDEGASGAMGFFNVHFSLASASITAMFS"
stream loop
5583..5605
5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.21; Mismatches 0;
Matches 25; Conservative 0; Indels 0; Gaps 0;
QY      1      cgatccctgcaggtgcaggtcagc 25
Db      221  CGATCCTGCTGACGTGCAGGTGACG 245

RESULT 7
A38251      5872 bp      DNA      PAT      05-MAR-1997
LOCUS      Sequence 6 from Patent WO9408023.
DEFINITION  A38251
ACCESSION  A38251
VERSION    A38251.1 GI:2294849
KEYWORDS
SOURCE
ORGANISM   Escherichia coli.
            Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE  1 (bases 1 to 5872)
AUTHORS   Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL   LONZA AG (CH)
            other publication PL 308301 950724
            other publication CA 2145400 940414
            other publication AU 4820293 940426
            other publication HU 71781 960228
            other publication SK 42095 951108
            other publication CZ 9500809 950913
            other publication FI 951547 950331
            other publication JP 85016947 960227.
FEATURES
Source
            location/Qualifiers
            1..5872
            /organism="Escherichia coli"
            /strain="DSM498"
            /db_xref="taxon:562"
            /clone="PB030A15-9"
            1141..1156
            /standard_name="BIOF RBS"
            1154..2308
            /gene="BIOF"
            1154..2308
            /gene="BIOF"
            /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
            /EC_number="2.3.1.47"
            /codon_start=1
            /transl_table=11
            /number=2
            /evidence=experimental
            /product="KAPA SYNTHASE"
            /protein_id="CAA02329.1"
            /db_xref="GI:2294850"
            /translation="MSMOEKINALLDARADALRRRYPVAGAGRWLVADDRQYLN
            SSNDYLGSHHPQILRAMOQGAEOFGISGSGSVSYVHQAEELEELAEWICYSRA
            LFIISGAANOAVIAAMAKEDRIADRLSHASILEASLSPSOLRRFANDVTLLAR
            LIAAPCGOAMVYEGVFSMDGSAFLAEIOOVYQHNGLMVDADHGTGIGDQNG
```

```
SCWLOKVPDLLVTFPGKRGVSGAIVCSSTVADYLLOFARHLIYSTMPAPAOAL
RASLAVIRSDSDGARREKRLALITRPAGQODLPFTLADSCSAIQPLIYDGNRRLQL
AEKIROQGCWYTAIRPTVPAGTARLRLTLAAHEMODIDRLLEVLHGNS"
RBS
/standard_name="BIOD RBS"
3030..3045
gene      3043..3753
CDS      /gene="BIOB"
            3043..3753
            /gene="BIOB"
            /standard_name="DETHIOBIOTIN SYNTHASE"
            /EC_number="6.3.3.3"
            /codon_start=1
            /transl_table=11
            /number=4
            /evidence=experimental
            /product="DTB SYNTHASE"
            /protein_id="CAA02330.1"
            /db_xref="GI:2294851"
            /translation="MSKRFVYGTDEYKTVASCALIOAKAAGRTAGKYPVSGS
            EKTPEGRMSDALAORNSLQDLYATVNPFAEPTSPHISAOEGRPIESLVMAG
            LRLBQADAVIVGACGMEPTLSDTFEADWVQEOPLVIVGKIGCINHAMLTA
            QVIOHAGLTLAGVANDVTPROKRAEYMTTLIRIPRLIGELIPWLAEENNAATGK
            YINLAFVDASTLGTFTSRL"
BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.21; Mismatches 0;
Matches 25; Conservative 0; Indels 0; Gaps 0;
QY      1      cgatccctgcaggtgcaggtcagc 25
Db      221  CGATCCTGCTGACGTGCAGGTGACG 245

RESULT 8
A93674      5872 bp      DNA      PAT      22-JAN-2000
LOCUS      Sequence 1 from Patent EP0798384.
DEFINITION  A93674
ACCESSION  A93674
VERSION    A93674.1 GI:6741862
KEYWORDS
SOURCE
ORGANISM   Escherichia coli.
            Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE  1 (bases 1 to 5872)
AUTHORS   Birch, O. and Brass, J.
TITLE     Biotechnological method of producing biotin
JOURNAL   LONZA AG (CH)
            location/Qualifiers
            1..5872
            /organism="Escherichia coli"
            /strain="DSM498"
            /db_xref="taxon:562"
            /clone="PB030A-15/9"
            1..96
            /function="PROMOTER PTAC"
            /evidence=experimental
            23..28
            /standard_name="PROMOTER PTAC"
            /evidence=experimental
            105..119
            /standard_name="BIOB RBS NO.9"
            /evidence=experimental
            117..1157
            /gene="BIOB"
            117..1157
            /gene="BIOB"
            /codon_start=1
```

```

/transl_table=11
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHRPRWTLISQVTELEFKPLDLLFEAQVHRQHPDPOVST
LSTITGACPEDECKYCPQSSRYKGLFEARLMEVQVLESARAKAAGSTFCGMAW
KNHEDMPLEOMVGVKAMGLEACMTLGTISESOAQLNAGIDYNNHLDTPGMAW
VGNITITTYOERLDTLEKRVAGIKVSGGIVGIGEVKCRAGILLDLANIPPPES
VFINLVKGTGLADNDVDVAFDFIRTAARIMPTSYRLSGRQNEOTQAMC
PMAGANSIEYCGKLTITTPNEDKDLQLEFKGLNLPQOTAVLAGDNEQOURLFOALMT
PDDEYVNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVKNQALAAFGRAAHYEQHADLQROSADALLAMLPQRYT
HYDAGCGPQMSHMRERHQAQVYALDLSPPMLVQARQKDAADHYLAGDIESTPLATA
TFDLAMSNLAOVQGNLSTALRELYRVRRGVVAFPTLVQSLPELHOMQAVDERP
HANRPLPDELEOSLNGVYOHNIQPTITMEDDALMSRLSGIKGATHLEGRDPRIL
TRSOLORIQLAMPQOQGRVPLTYHLFCVLAIRE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONANONATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLAPDRHRIHWPYTSPLPYVYVVSAGECELLSDGRR
LVDSMSWAAIHGVNHPOLNNAKSQIDAMSHVMFGTLTHAPLELCRKIVAMTPOP
LECVPLADSGVAVEYAKMAYLOQWQAKGEARQRLTFRNGYHGDTFGAMVSCDDNS
MSLWKGYLEPENLEFAPQSRMDGEWDERDMVGFALMAHHEITAIVIIPEYOGAG
GMRMYHEMLKTRIKICDREGILLIDETATGEGRTGLFACEHAELADITICLKL
TGTMTLSATLITREVAETISNGEACFPHGPTFMGNPLACAANAASIALIEGCMWO
OVADIEVOLREOLAPARDAEMVADVIGATIGVETTHPVNNAILOKFFPDGVNIRP
FGKLILMPYIILPQOLRLTAANRAVQDETFEQ"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDRQDRLPHRHVFNQMGYDGDNISPHLAMDVPAQIK
SFVVCYDPAFTGSGMHWVYNLPADTRVLPQFGSLVAMPDGVLOTRTFDQKTCG
VDGAAPFGEITHRIYFTVHALDIRIDVDEGASGAMVGFNVHFSIASITAMFS"
5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
1552..1695
BASE COUNT 1318 a 1552 c 1695 g 1307 t

```

```

ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgatccctgcaggtgcagtcagc 25
    |||
Db 221 CGATCCTGCTCAGGTGCAGTGCAGC 245

RESULT 9
A93679
LOCUS      A93679      5872 bp      DNA
DEFINITION Sequence 6 from Patent EP0798384.
ACCESSION A93679
VERSION   A93679.1 GI:6741867
KEYWORDS
SOURCE
ORGANISM  Escherichia coli.
          Bacteriophila coli
          Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
          Bacteriophila.
REFERENCE 1 (bases 1 to 5872)
AUTHORS  Birch O. and Brass J.
TITLE    Biotechnological method of producing biotin
JOURNAL  Patent: EP 0798384-A 6 01-OCT-1997;
          LONZA AG (CH)
FEATURES
source
Location/Qualifiers
1..5872
/organism="Escherichia coli"
/db_xref="taxon:562"
/clone="PB030A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONANONATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSWQEKINALDARRADALRRYPVAGAGRWLVADROYLNF
LSNDYLGISHPQIIRAMQOAGQFQIGSGSGSHVGSYVHQALEBELAEMIGYSRA
LLISFGFANQAVTAAMAKEDRIADRLSHLSLEAASPSQLRFRANDYTILAR
SCHWOKKPELVVTRKGFSGVSGAAVLCSTVADYLLDFARLLIYSTMPAQOAL
RASLAVTRSDGDRREKTLALITFRFAGVODLPFLADSCSAIQPLVGDNSRALQL
AEKIKQGCWVTAIRPPTVAGTARLRLTLTAHEMQDIDRILEVLHGNG"
3030..3045
/standard_name="BIOD RBS"
3043..3753
/gene="BIOD"
3043..3753
/gene="BIOD"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYEYVGTDEVEGKTVASCALLQAAKAGRYTRAGYKPVASGS
EKTPGELRNSDALQGRNSLQLDYATVAVPYFAPRTSPHISABGRPIESTIVSAG
LRLEQOAMVILEGAGGWEPTLSDTFPADVYTOBOLFYIIVGVCKLGCINHAATTA
OYIQAHGTLACVANDVTPPKRRAETWTTLRMIIPAVLGEITFWLAENPNAATGK

```

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 YINLAFVASTIGT.SRL
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.21;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgaatcctcgtcaggtcagtcagc 25
 ||||||||||||||||||
 Db 221 CGATCCTCGTCAGTGCAGTCAGC 245

RESULT 10
 ARI01809
 LOCUS ARI01809 5872 bp DNA PAT 14-FEB-2001
 DEFINITION Sequence 1 from patent US 6083712.
 ACCESSION ARI01809
 VERSION ARI01809.1 GI:12812607
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
 FEATURES
 source 1. 5872

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.21;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgaatcctcgtcaggtcagtcagc 25
 ||||||||||||||||||
 Db 221 CGATCCTCGTCAGTGCAGTCAGC 245

RESULT 11
 ARI01810
 LOCUS ARI01810 5872 bp DNA PAT 14-FEB-2001
 DEFINITION Sequence 6 from patent US 6083712.
 ACCESSION ARI01810
 VERSION ARI01810.1 GI:12812608
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
 FEATURES
 source 1. 5872

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.21;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgaatcctcgtcaggtcagtcagc 25
 ||||||||||||||||||
 Db 221 CGATCCTCGTCAGTGCAGTCAGC 245

RESULT 12
 AE000180
 LOCUS AE000180 11022 bp DNA BCT 01-DEC-2000
 DEFINITION Escherichia coli K12 MG1655 section 70 of 400 of the complete
 genome.
 ACCESSION AE000180 U000096
 VERSION AE000180.1 GI:1786988
 KEYWORDS
 SOURCE Escherichia coli K12.
 ORGANISM Escherichia coli K12.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

REFERENCE 1 (bases 1 to 11022)
 AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
 Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
 Gregor, J., Davis, N.W., Kirtpatrick, H.A., Goeden, M.A., Rose, D.J.,
 Mau, B. and Shao, Y.
 TITLE The complete genome sequence of Escherichia coli K-12
 JOURNAL Science 277 (5331), 1453-1474 (1997)
 MEDLINE 97426617
 PUBMED 9278503

REFERENCE 2 (bases 1 to 11022)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459

REFERENCE 3 (bases 1 to 11022)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459

REFERENCE 4 (bases 1 to 11022)
 AUTHORS Plunkett, G. III.
 TITLE Direct Submission
 JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

COMMENT
 This sequence was determined by the E. coli Genome Project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCHGR). The entire sequence was independently
 determined from E. coli K12 strain MG1655. Predicted open reading
 frames were determined using Genemark software, kindly supplied by
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
 have been correlated with genetic loci are being annotated with CG
 Site Nos., unique ID nos. for the genes in the E. coli Genetic
 Stock Center (CGSC) database at Yale University, kindly supplied by
 Mary Berlyn. A public version of the database is accessible
 (<http://cgsc.biology.yale.edu>). Annotation of the genome is an
 ongoing task whose goal is to make the genome sequence more useful
 by correlating it with other data. Comments to the authors are
 appreciated. Updated information will be available at the E. coli
 Genome Project's World Wide Web site
 (<http://www.genetics.wisc.edu>). ** The E. coli K12 sequence and
 its annotations are periodically updated; this is version M54. No
 sequence changes. Annotation updates: updated gene identifications
 and products; all new functional assignments courtesy of Monica
 Riley; added promoters, protein binding sites, and repeated
 sequences described in reference 1. The unique numeric identifiers
 beginning with a lowercase 'b' assigned to each gene (protein- or
 RNA-encoding) are now designated as gene synonyms instead of
 labels. This should allow them to be searched for in Entrez as gene
 names.

FEATURES
 source
 Location/Qualifiers
 1..11022
 /organism="Escherichia coli K12"
 /strain="K12"
 /sub_strain="MG1655"

```

repeat_region /db_xref="taxon:83333"
               <1..130
               /note="REP (repetitive extragenic palindromic) element;
               contains 4 REP sequences"
gene          complement(147..1430)
               /gene="ybhc"
               /note="b0772"
CDS           complement(147..1430)
               /gene="ybhc"
               /function="orf; Not classified"
               /note="f427; 98 pct identical to fragment ybhc_ECOLI
               SW:P46130 (300 aa) but contains 127 additional C-terminal
               residues"
               /codon_start=1
               /transl_table=1
               /product="putative pectinesterase"
               /protein_id="AAC73859.1"
               /db_xref="GI:1786989"
               /translation="MNEFSVRLALALAFVTLFACSTPPDPSDPTAAGTSRPI
               LSAEAFQNDQHFASLTGSAAMNSPILPQDPDFVGPACTGCTHTTIOAAYD
               AAIKRTKROYIAYMREYCTGYVPAAPGCTITLGGKEPITDKTGLSLDGMSTA
               DMRIQVNRKGYMPGKPAWYDSCQSRSDSICGLSAVFSQNGIQLQNLIENT
               LGDSVDAGNHPVALRTDGDQVQINNVLIRGOMTFEVTNSGVQNLLETNPOPTIVT
               NSYLEGDVIVSGGAVVFDFEYVNSRTOCEAYEAPATLSNIYGFVAVNSRFA
               AFGDGVQVQLGRSLVDANTNGOAVIRDSAINCEGFAPKPMADAVISNRPAGNGSVD
               DNDEIQRNLNDNTNRMENNRNGVSKVAEAKK"
promoter      complement(1507..1535)
               /note="factor Sigma70; predicted +1 start at 806574"
               /complement(1562..2058)
gene          /gene="ybhb"
               /note="b0773"
               /complement(1582..2058)
CDS           /gene="ybhb"
               /function="orf; Unknown"
               /note="f158; 99 pct identical to ybhb_ECOLI SW: P12994"
               /codon_start=1
               /transl_table=1
               /product="orf; hypothetical protein"
               /protein_id="AAC73860.1"
               /db_xref="GI:1786990"
               /translation="MKLISNDLRGDKLPHRHVFNMGYDGNISPLIANDVDPACTK
               SFVYCTDPDAPTSQGMHWVYVNLPAITVLPDGFSGSLVAMPDGLQTRIDFGKIG
               YDGAAPKRGTHRYIFTVHALDIERIDVDGASGAMVGFVHFHSLASASITAMFS"
protein_bind  complement(2108..2124)
               /note="central position to predicted promoter:85.5"
               /bound_moiety="Rnas predicted site"
               complement(2117..3406)
gene          /gene="bioA"
               /note="b0774"
               /complement(2117..3406)
CDS           /gene="bioA"
               /EC_number="2.6.1.62"
               /function="enzyme; Biosynthesis of cofactors, carriers:
               Biotin"
               /note="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
               /codon_start=1
               /transl_table=1
               /product="7,8-diaminopelargonic acid synthetase"
               /protein_id="AAC73861.1"
               /db_xref="GI:1786991"
               /translation="MTTDDLAFTQDRHMHPTSKTSLPYVYVASEGCELIISDGR
               LVDSMSWMAIHGYNHPOLNAAKSOIDMSVHVFGGTTHAIEICRKLVMTPQ
               LECVFLADSGSVAEYAKMAQLQYQAKGARQFLFRNGYGDFFGAMSVDDPNS
               MHSIMKGYLPENLEFAPQSRMDGEMDERMVFARLMAHRHIEIAAVIIEPIVQAG
               GMRMYHEMLKRIKICDRGILLIADIELTGGRGKILFACCHAEIADPIDICGAL
               TGGMTLSATLTREVAETISNGEAGCPMFKNPGLACAAANSIALIIEGDMQ
               OVADIEVQLREQLAPADAMVADRYLGLGIVETTHFVNNALOKFEVEGCVWIRP
               FGLIIMPYITLIPQDLRLTAANNAVODETFEQ"
promoter      complement(2193..2221)
               /gene="bioA"
               /note="factor Sigma70; predicted +1 start at 807260"
               3411..3450
               /note="central position to bioB promoter: -20"

```

```

protein_bind  /bound_moiety="BioB documented site"
               3411..3450
               /note="central position to predicted promoter: -20; genetic
               evidence for the site"
protein_bind  /bound_moiety="BioB predicted site"
               complement(3411..3450)
               /note="central position to predicted promoter:50"
               /bound_moiety="BioB predicted site"
               complement(3411..3450)
               /note="central position to bioA promoter:50"
               /bound_moiety="BioB documented site"
               3413..3441
               /note="factor Sigma70; promoter bioB; documented +1 at
               808525"
               complement(3447..3473)
               /note="factor Sigma70; promoter bioA; documented +1 at
               808515"
gene          3493..4533
               /gene="bioB"
               /note="b0775"
               3493..4533
               /gene="bioB"
               /EC_number="2.8.1.-"
               /function="enzyme; Biosynthesis of cofactors, carriers:
               Biotin"
               /note="0346; 99 pct identical to BIOB_ECOLI SW: P12996"
               /codon_start=1
               /transl_table=1
               /product="biotin synthesis, sulfur insertion?"
               /protein_id="AAC73862.1"
               /db_xref="GI:1786992"
               /translation="MAHRPRWLSQVTELFEXPIIDLFEAQVHROHFPDROQVST
               LSIKRGACBEDKTCPOSSRYKTGLEBRLEWVEQVLESARAKAAAGSTRFCGAAM
               KNHEDMPYLEQVQVQKAMGLACMTLGTISEQAOQLANAGIDYNNHLDTPES
               YGNITTRTYQERLDITLEKVRDAGIKVSGGIVGEGVKRAGILLOLANIPPEP
               VPIINMLVKKGPTIADNDVDADFRTITAVARLIMPTSVYLSGREGMNEOQAMC
               FMAGNSIFPGCKLITTPPEEDKDLQIFRKGLNPGQTAVVLAGDNEQQLLEQALMT
               PPTDERYNAAL"
               4530..5684
               /gene="bioF"
               /note="b0776"
               4530..5684
               /gene="bioF"
               /EC_number="2.3.1.47"
               /function="enzyme; Biosynthesis of cofactors, carriers:
               Biotin"
               /note="0384; 100 pct identical to BIOF_ECOLI SW: P12998"
               /codon_start=1
               /transl_table=1
               /product="8-amino-7-oxononanoate synthase"
               /protein_id="AAC73863.1"
               /db_xref="GI:1786993"
               /translation="MSWQEKINAAIDARRADALRRRYVPAQAGRWLVADROYLNF
               SSNDYLGSHHPQIRAWQGAEQFGISGSGSHVSYVYHQALEELAEMLGYSRA
               LFIISGFANQAVIANAMAKEDRIADRLSHASILEASLPSOLRRRAHNDVYHAR
               LLASPCQGMVYVEGVFSMDGSDAPLAEITQGVQOHNQGMVMDAHSOTGVIGGQGG
               SCMLQKVBELLVYTFGKFGVSGAAVLCSTVDVYLQIFARHLIYISMPAPAQAL"
Query Match 100.0%; Score 25; DB 1; Length 11022;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 cgaatcctcgcaagtcagtcagc 25
    |||||||||||||||||||
Db 3597 cgaatcctcgcaagtcagtcagc 3621

```

```

RESULT 13
LOCUS AE005258 13501 bp DNA BCT 21-MAR-2001
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
of 155.
ACCESSION AE005258 AE005174

```

```

VERSION      AE005258.1 GI:12513751
KEYWORDS
SOURCE       Escherichia coli O157:H7 EDL933.
ORGANISM     Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
              Escherichia.
REFERENCE    1 (bases 1 to 13501)
AUTHORS      Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
              Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
              Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
              Grobbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
              Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
              Welch,R.A. and Blattner,F.R.
              Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
              Nature 409 (6819), 529-533 (2001)
TITLE        JOURNAL MEDLINE
PUBMED       21074935
REFERENCE    2 (bases 1 to 13501)
AUTHORS      Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
              Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
              Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
              Grobbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
              Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
              Welch,R.A. and Blattner,F.R.
              Direct Submission
              Submitted (22-OCT-2000) Laboratory of Genetics, University of
              Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
              Location/Qualifiers
              1..13501
              /organism="Escherichia coli O157:H7 EDL933"
              /strain="EDL933"
              /serotype="O157:H7"
              /db_xref="taxon:155864"
              /note="enterohaemorrhagic"
              <1..7576
              /note="O-island #36: Region of the EDL933 chromosome not
              homologous to E. coli K-12 MG1655; Cryptic prophage
              CP-933K; includes one copy of the 13 bp direct repeat that
              flanks the prophage"
              66..665
              /gene="lomK"
              /note="lomK"
              /function="putative membrane; Other or unknown (Phage or
              Prophage Related)"
              /note="Residues 1 to 199 of 199 are 72.86 pct identical to
              residues 1 to 199 of 199 from Genpept 118 :
              g117532789|gb|AAFG3231.1|AF151091_2 (AF151091) lom
              [prophage P-ElDA]"
              /codon_start=1
              /transl_table=11
              /product="putative outer membrane protein of prophage
              CP-933K"
              /protein_id="AAG55137.1"
              /db_xref="GI:12513752"
              /translation="MKKVCALIIISAICLAIVSPKASEHOSTLSAGTLASTDAPC
              SDDLNGINVKRYEFTDTGLITFSYANAEDEOKTHYSTDTMHEDYVKNRFSTWAG
              PSVIVNEMFSAYAMAGVAYSRVSTFGSDYFRYDNKRKTHDVLTSDDARYSTSLAW
              GAGVQFPTSEVAVDAVEYKSGSGDMRTDGFIVGVYKF"
              719..2041
              /gene="20982"
              /note="20982"
              /function="putative structure; Structural component (Phage
              or Prophage Related)"
              /note="Residues 164 to 440 of 440 are 68.79 pct identical
              to residues 381 to 645 of 645 from Genpept 118 :
              g114585436|gb|AAD25464.1|AF125520.59 (AF125520) putative
              tail fiber protein [Bacteriophage 933W]"
              /codon_start=1
              /transl_table=11
              /product="putative tail component of prophage CP-933K"

```

```

gene         /protein_id="AAG55138.1"
              /db_xref="GI:12513753"
              /translation="MGMAAVDVISVLLKDGAKPIONCTIQLKRRNSTVYVNVASE
              NPDEGRYSMDVEYGYSTVLLVGEFPSSHAGTISYEDSQPTLNDPFGKAMEDAR
              PEALRRPEQVMEARAHAEKKNAGAEETISARNAISAKASAKANADTSADSEADSE
              SAROAFSAASAKKSEFASSSSSAQAQSESIQSSTYDAELSKTAESNAGNAADA
              TTSSEKASASASQAOSRISAEADVNIIPVYVGPFGPKGPGAGPGGPGGADGA
              KCPGPKATGTEGSGDGTGPGAGPGKSGRGETGLGNAGPGPGKPGGTGAAGPA
              GPGCGKGTGAAGVYATGATGPGPGKPGCGEIOIPRIGPMTITETNSGMFPGTDGLI
              TGLTFLDPKDATVOYGMFHLQVRFSGGPMQDVKGLDEVSDFGRGE"
              2004..2312
              /gene="20984"
              /note="20984"
              /function="orf; Other or unknown (Phage or Prophage
              Related)"
              /note="Residues 1 to 102 of 102 are 98.03 pct identical to
              residues 28 to 129 of 129 from Genpept 118 :
              g114585437|gb|AAD25465.1|AF125520.60 (AF125520)
              hypothetical protein [Bacteriophage 933W]"
              /codon_start=1
              /transl_table=11
              /product="unknown protein encoded by prophage CP-933K"
              /protein_id="AAG55139.1"
              /db_xref="GI:12513754"
              /translation="MKMAVIOALENDKMLRLKIMQSLCGCGKHDCENGSRSLTAOLR
              LCPADLDESDGILPEODRVTIVYIILADKKQIQCVVRPLQILRADGTWENIGMK"
              2489..3469
              /gene="20985"
              /note="20985"
              /function="orf; Other or unknown (Phage or Prophage
              Related)"
              /note="Residues 13 to 326 of 326 are 55.73 pct identical
              to residues 18 to 331 of 336 from Genpept 118 :
              g116960367|gb|AAFG3527.1 (AF170176) hypothetical protein
              predicted by Glimmer [Salmonella typhimurium LT2]"
              /codon_start=1
              /transl_table=11
              /product="unknown protein encoded by prophage CP-933K"
              /protein_id="AAG55140.1"
              /db_xref="GI:12513755"
              /translation="MLSPIRTTFHNSVNIQVSSPCQTVSPFAGREYELKVIDEKPILF
              OMPEPNPERKKEVPIVNTKQHPYLDVNTNARIESRMIGIFVDCGFSVNOKTAES
              KLERDPNNVITIRYEDVDFSMYDRKLSIDYHDIICEORLRTEDKDEYLLMLLEKLR
              EISKADSLISMTAKKRNNAWDFPNNALAKAGIFPCTVYTKRNGHISFEGCIYLD
              MDMLTGLKGTIYAPDGISMHVDRNDSVNIENSALTYNRSNHPALLGSLSPMNSKVD
              AHBYDGLKGVKKRYFNPLNHNHFCDFTEFNHPIIMNTSQTTCSSW"
              3530..4522
              /gene="20986"
              /note="20986"
              /function="orf; Other or unknown (Phage or Prophage
              Related)"
              /note="No significant matches"
              /codon_start=1
              /transl_table=11
              /product="unknown protein encoded by prophage CP-933K"
              /protein_id="AAG55141.1"
              /db_xref="GI:12513756"
              /translation="MKIPISLQSNFNSAPAGSAPISAPPAKRAENAVDYLDIGKRPL
              SAADLSNVESVIRAVDSRSRLIDHTYDMIGNVTLADLSQSFRDRAVSTGHNKX
              VHICICYRNEVELNESVYKIDDIQSLCNELYEDVQDEPFPICEGENDNEDEPY
              VSFVAPDTSYEMPSWQEGILHEIILHNHYSGLSDPSGSDNIEIGPTIELARVADQLG
              MSVPDQGVGAPEPEEAHLRLRNALROAMRHEENRAFPRLGTISRYEASDPT
              ESKVSNIGYGFIOQHPLALINDMLQDANQOLYHGAPVYTFEDVDKHNQO"
              5350..6231
              /gene="20989"
              /note="20989"
              /function="orf; Other or unknown (Phage or Prophage
              Related)"

```

```

/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AG5142.1"
/db_xref="GI:12513757"
/translation="MISPYSEVHIGCSMNSLTNRLTPDNRVLSSRDAVHSDGACV
KGNRTYRVATDMKFCVTRSHSGCFNLHLRIGMPKGETSRKTEVMANSPYSAM
ERGIYHSNRDLPVDVAPPELPSVDNRISVPCNVIGKGNAYVEDADATKVLK
FTTSQNEEVTSEVRCFNOYAGASAEITNNNDIGIRMDKINGSLNLSLPAO
AEHAIYMDPELEOKGILFVDTTETNVLVDRAKNEFNDISSYVSDRSSESQIMQ
SYHGKODLISVLSKI"
/gene="20990"
/length=7160
/size=6462
/gene="20990"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AG5143.1"
/db_xref="GI:12513758"
/translation="MRPTSLNVLVHSSRSSMSDDIESLVASSVOMIKNNPOLRF
QCTDNLITQGLEALDKIGSTETGRVLLNIESIRLSTETVTHNSSRLGYMAHD
IDENHRTGSDHFCHNNAVEYPCGEGIVGEGHATVFEHLHAFHNLNGERLAVES
SRASOKYSPLLLEARTVGLAFSEFEVLSSEKFEHETGMPRTSYPSXALHHDNT
VSLGFOQVRLHPL"
/gene="20992"
/length=7668
/size=6462
/gene="20992"
/function="orf; Unknown function"
/note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MG1655; R0773"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AG5144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLRDGDKLPHRVFNGMGYDGNISPHLAMDVPAGTK
SFVYCYDDAPGSGMMWVYVNLPAIDTRVLPQGGSLVAMPQSVLQTRIDFGKTC
YGAAPKGETHRTYFTVALDVERIDVEGASGAMVGNVHFHSLASASITAMFS"
/gene="20993"
/length=8203
/size=6462
/gene="20993"
/complement(8203..9492)

CDS
Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgaacctcgtcaagtcagtcagc 25
Dh 9683 CGATCCTCGTCAGCTCAGCTCAGC 9707

RESULT 14
AP002553 LOCUS AP002553 297816 bp DNA BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION AP002553 BA000007
VERSION AP002553.1 GI:13360211
KEYWORDS
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub-strain:RIMD 0509952)
DNA.
ORGANISM Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (sites)

/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AG5142.1"
/db_xref="GI:12513757"
/translation="MISPYSEVHIGCSMNSLTNRLTPDNRVLSSRDAVHSDGACV
KGNRTYRVATDMKFCVTRSHSGCFNLHLRIGMPKGETSRKTEVMANSPYSAM
ERGIYHSNRDLPVDVAPPELPSVDNRISVPCNVIGKGNAYVEDADATKVLK
FTTSQNEEVTSEVRCFNOYAGASAEITNNNDIGIRMDKINGSLNLSLPAO
AEHAIYMDPELEOKGILFVDTTETNVLVDRAKNEFNDISSYVSDRSSESQIMQ
SYHGKODLISVLSKI"
/gene="20990"
/length=7160
/size=6462
/gene="20990"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AG5143.1"
/db_xref="GI:12513758"
/translation="MRPTSLNVLVHSSRSSMSDDIESLVASSVOMIKNNPOLRF
QCTDNLITQGLEALDKIGSTETGRVLLNIESIRLSTETVTHNSSRLGYMAHD
IDENHRTGSDHFCHNNAVEYPCGEGIVGEGHATVFEHLHAFHNLNGERLAVES
SRASOKYSPLLLEARTVGLAFSEFEVLSSEKFEHETGMPRTSYPSXALHHDNT
VSLGFOQVRLHPL"
/gene="20992"
/length=7668
/size=6462
/gene="20992"
/function="orf; Unknown function"
/note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MG1655; R0773"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AG5144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLRDGDKLPHRVFNGMGYDGNISPHLAMDVPAGTK
SFVYCYDDAPGSGMMWVYVNLPAIDTRVLPQGGSLVAMPQSVLQTRIDFGKTC
YGAAPKGETHRTYFTVALDVERIDVEGASGAMVGNVHFHSLASASITAMFS"
/gene="20993"
/length=8203
/size=6462
/gene="20993"
/complement(8203..9492)

CDS
Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgaacctcgtcaagtcagtcagc 25
Dh 9683 CGATCCTCGTCAGCTCAGCTCAGC 9707

RESULT 14
AP002553 LOCUS AP002553 297816 bp DNA BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION AP002553 BA000007
VERSION AP002553.1 GI:13360211
KEYWORDS
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub-strain:RIMD 0509952)
DNA.
ORGANISM Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (sites)

AUTHORS
Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C. H., Kimura, S.,
Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T.,
Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
Sasakawa, C. and Shinagawa, H.
TITLE
Complete nucleotide sequence of the prophage v72-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
20198780
JOURNAL
MEDLINE
REFERENCE
AUTHORS
Kurokawa, K., Murata, T., Nakayama, K., Kuhara, S., Hattori, M.,
Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and
Hayashi, T.
TITLE
Comparative analysis of the whole set of RNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356
JOURNAL
MEDLINE
REFERENCE
AUTHORS
Yokoyama, K., Makino, K., Kubota, Y., Matanabe, M., Kimura, S.,
Yutsudo, C. H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and
Shinagawa, H.
TITLE
Complete nucleotide sequence of the prophage v71-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Genes 258 (1-2), 127-139 (2000)
20564182
JOURNAL
MEDLINE
REFERENCE
AUTHORS
Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T.,
Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C.,
Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and
Shinagawa, H.
TITLE
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
21156231
JOURNAL
MEDLINE
REFERENCE
AUTHORS
Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and
Hayashi, T.
TITLE
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail: ken@gen-info.osaka-u.ac.jp,
url: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,
Fax: 81-6-6879-2047)
5 (bases 1 to 297816)

COMMENT
FEATURES
location/Qualifiers
1..297816
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
79..1245
/gene="ECs0753"
79..1245
/gene="ECs0753"
/note="similar to SUCC-ECOLI g11786948 percent identity
100 in 388 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="succinyl-CoA synthetase beta subunit"
/protein_id="BAB34176.1"
/db_xref="GI:13360212"
/translation="NNLHXYQAKOLFARYGLPAPVGYACTPREAEEAASKIGAGPW
VKQVHAGCRGAGAGVKNVSKEDIRAFAPENLGRVLYOTDAGGPNOLIVAAAT
DIAKELYLGAVDVRSRRVFNMASTEGVEITKVAEENPHILHKXALDPLGPMYOG
RELAPFLGLEGLVOOFTKIFMGALDPIELERLALIEINPLVITRQGDICDGLGA
DGNALFRODILREMDQSOEDPREQAAQWELNVALDGNIGCVNAGLAWGMDIIV
KIHGEPANLFDVGGATKREVAFKIILSDQKVALVNTFGIVGICDLIADGIIIG
AAVEGVNPPVVRLEGNNAELGAKRLADSGLIIIAKGLTDAQAQVAAVEGK"

```

```

gene
1245. .2114
/gene="ECS0754"
CDS
1245. .2114
/gene="ECS0754"
/notes="similar to SUCD_ECOLI g111786949 percent identity
100 in 289 aa (conserved in E.coli K-12)"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/translation="MSTLIDKNTKVCIGGTTGSGGTFTHSEQATAYCRKMGVTPCKG
GTHLGLPVTNVAEVAATGATASVIVPAPCKOSILEAIDAGIKLITITTEGIFT
LDMITVYKILDEACVPMIGPNCPCVITPGCEKIGIOPGTHKPKGVIVRSQGLIYE
AVKOTDYEGGOSTGCVGIGDPGPGSNFIDLEMFEDPQTEAIVMIGELGSGAESEA
AAVTEHVTPVGVYINGVIAPKGKRNGHAGIAGCKGTADEKFALEAAVAKTVNS
LADIGALKTYLK"
2781. .3686
/gene="ECS0755"
2781. .3686
/gene="ECS0755"
/notes="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
transcription activator (gcv operon activator) -
Escherichia coli g114170431sp|P32064|GCV_A_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"
/translation="MRGKIPKTELLVFEVVAHREHYTRAEEALATQSAVROVSAL
EEFRTPLFSKSKRIPLNDGKYLGIVKETLNLKLEEDTMTIMTQVYVIELAVN
PPTFTHMLIPNLHETTKLPDIIVNHSLANGNEVDAYIMREKFCAPMAEVEY
LPEEFLIPVCGSLMAGSDOKLSVAEILTEPLLHOSRTITGMEEMFALSGVSSPLVN
NGPRDLISMLIAVRSNLGYALLPRFAIQHDDSDGVIPICDVPIRGNFMITWCE
EKSDSPHLDQFREMILAKSVYPOEM"
complement(3720. .4322)
/gene="ECS0756"
complement(3720. .4322)
/gene="ECS0756"
/notes="probable cob(I)alamin adenosyltransferase, similar
to cob(I)alamin adenosyltransferases (corrinoid
adenosyltransferases) e.g. Escherichia coli
g11151481sp|P13040|BTRUR_ECOLI percent identity 67 in 200
aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="putative cob(I)alamin adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/translation="MEARISTERNHVROOQKLEQVDTVAALIEKKGLIYFTGNGK
KSTAFGTAVRAHGKTYGVAOYIKGMDNGEYMLDPLGVEFHMGFTWETONR
QADIDAKRVMSKSKRLADKRYDLVDELTYMLAYHLDTEEVYIASLQNNPAQGV
IYTGRCQSQIILKMDTVSEIRPVKHAHDNGIQAPGIDM"
complement(4332. .5984)
/gene="ECS0757"
complement(4332. .5984)
/gene="ECS0757"
/notes="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
(fumarate) - Escherichia coli
g111205981sp|P00923|FUMA_ECOLI percent identity 68 in 545
aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"

```

```

/transl_table=1
/evidence=not_experimental
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translation="MKKISLITWMLIALVIGMLTGVNLNTASPETAKLYAOEISIFT
TIFRLIKMILAVSTIVYGIAGKMDKALCRISPLFLFICSLISIALGLITV
NFMPTGGINFVAHGAETTCVAAEPTLITSHAFPIYDAMAHNETIIOYVESI
FLUCSLTAIOEKGSATVHALDSLAAHMLKITEGVMLFAPLTPAASAIARGLAVM
VSAIGTFMGEYFPMILLWVLLIGLAVVYGPCLIRLRLSPALLATFTSSSEAF
GTLEKEOFEGSPKIASFVLPIDGVSFNLGSMAYCFVAFVIAQACNHLISIGEOTM
LLIEMLTSGMAGVPRASVVIATVATNORIPAGIILLMGVDFPLDMGRSATVNSN
AMCAAMVSRWEGFEGCGKALKNESVALP"
complement(7533. .7850)
/gene="ECS0759"
complement(7533. .7850)
/gene="ECS0759"
/notes="unknown"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"

```

Query Match 100.0%; Score 25; DB 2; Length 297816;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 cgatcctcgtcagtgatgcagtcagc 25
DB 92415 CGATCCTCGTCAGTGCAGTCAGC 92439

```

RESULT 15
AF250776
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE

AF250776 5526 bp DNA BCT 31-JAN-2001
uncultured bacterium pCOSH2 hypotheical 17.1 kDa protein in
modc-bioa intergenic region, DAPA-aminotransferase Bioa (bioa),
biotin synthase Bioa (bioa), KAPA synthetase Biof (biof), and
biotin biosynthesis protein Bioa (bioa) genes, complete cds, and
dethiolobiotin synthetase Bioa (bioa) gene, partial cds.
AF250776
AF250776.1 GI:12620124
uncultured bacterium pCOSH2.
uncultured bacterium pCOSH2
Bacteria; environmental samples.
1 (bases 1 to 5526)
Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streitz,W.R.
Direct cloning from enrichment cultures, a reliable strategy for
isolation of complete operons and genes from microbial consortia
Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
20575196

PUBMED 11133432
 REFERENCE 2 (bases 1 to 5526)
 AUTHORS Entcheva, P., Liebl, W. and Streitz, W.R.
 TITLE Direct Submission
 JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet Goettingen, Griesbachstr. 8, Goettingen 37077, Germany
 FEATURES
 source
 1..5526
 /organism="uncultured bacterium pcosH2"
 /db_xref="taxon:143797"
 /clone="pcosH2"
 /note="unknown organism, cosmid clone derived from environmental consortium"
 complement(52..528)
 /note="ORF1"
 /codon_start=1
 /transl_table=11
 /product="hypothetical 17.1 kDa protein in modC-bioA intergenic region"
 /protein_id="AAG60577.1"
 /db_xref="GI:12620125"
 /translation="MKLISNDLRDGLKPHRHVNGMGYGDNI SPLANDVPAGTK SEVVCYDDPDAPITSGGMMHWYVLPADIRVLPQSGSLVAMPDGLVQRTFGKAG YDGAAPKRGTHRTFTVHALDIRIDVDEGASGAMGVFNHFSLSASITAMFS"
 complement(587..1876)
 /gene="bioA"
 complement(587..1876)
 /note="7,8-diaminopelargonic acid synthetase-aminotransferase"
 /codon_start=1
 /transl_table=11
 /product="DAPA-aminotransferase bioA"
 /protein_id="AAG60578.1"
 /db_xref="GI:12620126"
 /translation="MTTDDLAFDQRHIMPTYSKTSPLPYVVSAGCELLISDGR LVDSMSWMAIHCYTHNPOLAAKSOIDMSHVMTGGITHAIELCRLVAMTPOP LKCVFLADSSVAEFAVAMKMAIYQWAKGARGORFLFRNGYHGDGMSVCDPDS MSLKMGKYLPELNFAPAPQSRMDGEMDERMVFARLMAHREHIAVITEPIVQAG GMMVYHPEMLKRIKRICDREGILLADELLEGRGKILFACGHAETIADICIGAL TGGMTLSATLITREVAETISNGEAGFMHGPMPKMLACAAANSLALBESGMOO OVADIEVOLREOLAPADAMVADVRLGIGVETTHPYVMAALOKFEVEGVWIRP FGLTIYMPYITILPQOLRLTAANNAVODETFQO"
 1963..3003
 /gene="bioB"
 1963..3003
 /note="bioB"
 /codon_start=1
 /transl_table=11
 /product="biotin synthase bioB"
 /protein_id="AAG60579.1"
 /db_xref="GI:12620127"
 /translation="MAHRRPWLSCVTELEFEKPLDLFEAQOVIHROHDFPROVOST LLSIKTGACPEDECKYPOSSRYKTGLEARIMEVQVLESARKAKAASGTRCMANM KNENRMDPELEOMVOGVDLGEACMTGCTISEQAOLANGLDYTHNIDPTSEF YGNITRTTYOERLDLEKRVAGIGVCGGIVGAEYTKDRAGLLLQLANLPTPEES VPRNMLYKKGCTPLADNDVDADPRTITAVARIMPTYSVRLSAGREOMQOTQAMC PMAGASIFYCGKLTTPNEEDKDLQIFRKGLNPOQRAVLADGNEQOQRLQALMT PTDEYVNAAL"
 3000..4154
 /gene="bioF"
 3000..4154
 /note="8-amino-7-oxononanoate synthase"
 /codon_start=1
 /transl_table=11
 /product="KAPA synthetase bioF"
 /protein_id="AAG60580.1"
 /db_xref="GI:12620128"
 /translation="MSMOCKTNAALDARRADALRRRYPAVAGAGRWLVADROYLNF SSNDYLGISHPOITIRAMQOGEFGISGSGSHVGYVAQALREBLALWLGYSRA LFIISGANQAVIAMAAMAKEDRIADRLSHASLLEASLSPSOARRPAHDVTHLAR LLASPCGOOLVYVEGVFSMDGDSAPLAEIQOVTQOHNGMWLVADAHGTGYISGGRG

gene
 CDS
 4141..4896
 /gene="bioc"
 4141..4896
 /note="reaction step prior to synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=11
 /product="biotin biosynthesis protein Bioc"
 /protein_id="AAG60581.1"
 /db_xref="GI:12620129"
 /translation="MATVKNQALIAAFGRNAALYEQHADLQROSDAVLLAMPERKYT HVLDACCGGEMWT RHWREHRAQVTAIDLSPMLVOAROKDADHYLAGIESTPLATA TEDIANSLAYOVCGNLSTALRELYRVSKGVAPETLVQSGSLPELHOAMQAVDERP HANRFLPDETEIOSLNGVYOHIOPTILMPDALSAMRSKLGITATHLHESRDRIL TRSOLOQLAMPQOQGRPLVYHFLGYIAKE"
 4883..5526
 /gene="biOD"
 4883..5526
 /note="DHB synthetase"
 /codon_start=1
 /transl_table=11
 /product="delhibiotin synthetase bioD"
 /protein_id="AAG60582.1"
 /db_xref="GI:12620130"
 /translation="MLVSKRYFVTGTDTEVNGTVASCALLQAAKAGYRTAGYPVYS GSEKTPGGRNSDALALQNSSLOLDVTPYPAFTSPHIIISAQGRIESVMS SGLRALQOADMVIVEGAGCMPTPSDPTFRADYTOQLVILVQKLCINBAML TNOAIOHAGLITACGAVADVTPGKRHAEYITITLIRMLPAVLDELIPWLAENPE"
 1507 g 1178 t
 BASE COUNT 1274 a 1507 c 1567 g 1178 t
 ORIGIN
 Query Match 93.6%; Score 23.4; DB 2; Length 5526;
 Best local Similarity 96.0%; Pred. No. 1.2;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 cgaatcctgcagglcagcagcagc 25
 ||||||||||||| |||||||||
 Db 2067 CGATCCTCGTCAGGTCAGGCAGC 2091
 RESULT 16
 SMABIO
 LOCUS SMABIO 7215 bp DNA BCT 04-FEB-1999
 DEFINITION S. marcescens biotin operon, complete sequence.
 ACCESSION D17468
 VERSION D17468.1 GI:402530
 KEYWORDS 7,8-diamino-pelargonic acid aminotransferase;
 7-keto-8-aminopelargonic acid synthetase; DAPA aminotransferase;
 DHB synthetase; KAPA synthetase; biotin operon; biotin synthetase;
 decthiobiotin synthetase.
 SOURCE
 ORGANISM Serratia marcescens (strain:5r41) DNA.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Serratia.
 1 (bases 1 to 7215)
 REFERENCE
 AUTHORS Sakurai, N.
 TITLE Direct Submission
 JOURNAL Submitted (25-AUG-1993) to the DDBJ/EMBL/Genbank databases. Naoki Sakurai, Tanabe Seiyaku Co., Ltd., Res Lab of Applied Biochemistry; 2-50, Kawagishi-2-chome, Toda-shi, Saitama 335, Japan (E-mail:grc0110@niftyserve.or.jp, nsakurai@dcbj.nig.ac.jp, Tel:048-433-2545, Fax:048-433-2540)
 REFERENCE 2 (bases 1 to 7215)
 AUTHORS Sakurai, N., Imai, Y., Akatsuka, H., Kawai, F., Komatsubara, S. and Tosa, T.
 TITLE Complete nucleotide sequence of biotin operon of Serratia marcescens
 JOURNAL Unpublished (1993)


```

/gene="b10A"
/codon_start=1
/transl_table=11
/product="B10A-like protein"
/protein_id="AAG60558.1"
/db_xref="GI:12620098"
/translation="MSVTDSDLEFDQRIHWPYTSMSHPDPCYVEAASGVEIQADG
RALDGMSSMAAIIHGTNHPQNLAIQQLKMSHVMFGITHEANELCKRYAMTP
EALQCVPLADSGSAVEYSIKMALQYQWARGERQRIITLGRGHGPTFGAMSYCDPD
NSMISLYGYLAFLHFAPOQCFDQADENDIAPFALLBQHAGEVAAYLLEFVPG
AGMRILHPTLYLERELCDRYOVILLADEIATGEGRTGKLFACHEHAEPYDILCTCK
ALTGVTMLSATITTRHYAETISNCAAGCEHPMGNPLACAVANASISLAENK
QQQYAAIETOLKQOLLPLKMPKVADVAVLGAIGVEKREPVMAQLQGRVGVNI"
1265..2305
/gene="b10B"
1265..2305
/gene="b10B"
/codon_start=1
/transl_table=11
/product="B10B-like protein"
/protein_id="AAG60559.1"
/db_xref="GI:12620099"
/translation="MMADRIHWTYGLAQTLLFKPLLELFEAQTVHRHROHPPROYVS
TLISIKGACPECKYCKPOSSRYKTELEPELMQVBOYLISAKAKANGSTRQMGAP
MKNHERMDPYLOOMOGCYKAMGNETCTGCTDGSQAEKAEGLDYVHNLDTSPE
FYGNITTRISQERLTLGKVRGAGIVCSGGIVGLGETVYXRAGILVOIANDLTPPE
SVPLNMLKVRAGTPLANDVDVDPDFRTIYAVARIMPPSSVYRLSAGREOMECOTAM
CEMAGANSIFYGCKLITTPNPEEDKDLQLRKILGILNPQOTETEHGNOQOEVLARIM
QADTAFFYNAL"
2305..3456
/gene="b10C"
2305..3456
/gene="b10C"
/feature="7-keto-8-aminopelargonic acid synthetase; similar
to B10F"
/codon_start=1
/transl_table=11
/product="KAPA-synthetase"
/protein_id="AAG60556.1"
/db_xref="GI:12620096"
/translation="MSWQRIEQLERXSOAAYRHRVLKGGNGRYLQNDQRYLNF
SGNDYGLSONAOVIAIMOQGAQOYGVSGSGHVTGCTAHOALRQOALQOMGYPS
LLEFISGAANOALIALMOGEDRIADRLSHASLLPAAHSPALRRFRINQOALAD
LLKPCSGQTLVATEGVSMGDSAPLAELRYLRDAGAWLVDDAHGIGVGEORG
SCWQGVREPLLVTEFGKARGLSGAAVLCDESTAEYILQPARHLITSTAMPRAQACAL
QALHCHIOGDELRTLNTNIOFRQGANLSTITESTAIOTPLIVGDNORALDLAQ
HIRGQGLMSAIRPPYPPGAGRLRITLTAHQPDIDRLLEVLDHVS"
3440..>4000
/gene="b10C"
3440..>4000
/gene="b10C"
/feature="similar to B10C; probably reaction step prior to
synthesis of pimeloyl-CoA"
/codon_start=1
/transl_table=11
/product="biotin biosynthesis protein B10C"
/protein_id="AAG60557.1"
/db_xref="GI:12620097"
/translation="MTSAFEQVNVKQAVASAFSAAGYAAVQLQDVGERRLQKOE
HOGEOILLDAGCGTGYFSRRKRELKQVTLADLAPGMLARQOQADHVLQDTEVP
LGDAAVDISFSSLYVOWCSDLPRAIAELRVTGPGVILLFTLAEGLHEDLDAMQOV
DGRHVNDELPLAQIEACAGYRHHQ"
BASE COUNT 794 a 1164 c 1236 g 802 t 4 others
ORIGIN
Query Match 83.2%; Score 20.8; DB 2; Length 4000;
Best Local Similarity 91.7%; Pred. No. 21;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
2 gatcctcgtcaggtgcagtcagc 25
||||| ||||||| ||||| |||

```

```

Db 1373 GATCCGCGTCAGTGTGACGTGAC 1396
RESULT 18
EBHIOTOP2
LOCUS EBHIOTOP2 1041 bp DNA BCT 04-NOV-1996
DEFINITION Erwinia herbicola biotin synthetase (b10B) gene, complete cds.
ACCESSION U38648
VERSION 038648.1 GI:1228112
KEYWORDS 2 of 2
SEGMENT Erwinia herbicola strain-Eh10.
SOURCE Pantoea agglomerans
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Pantoea.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Wu,C.H., Chen,H.Y. and Shuan,D.
TITLE Isolation and characterization of the Erwinia herbicola b10 operon
and the sequences of the b10A and b10B genes
JOURNAL Gene 174 (2), 251-258 (1996)
MEDLINE 97045821
REFERENCE 2 (bases 1 to 1041)
AUTHORS Shuan,D.
TITLE Direct Submission
JOURNAL Submitted (14-OCT-1995) David Shuan. Biology, Natl. Sun Yat-Sen
University, Kaohsiung, Taiwan, ROC
FEATURES
source
1..1041
location/Qualifiers
1..1041
organism="Pantoea agglomerans"
strain="Eh10"
db_xref="taxon:549"
1..1041
/gene="b10B"
1..1041
/gene="b10B"
/codon_start=1
/transl_table=11
/product="biotin synthetase"
/protein_id="AAG4534.1"
/db_xref="GI:1228114"
/translation="MAHARWTLISQVTELFPRKPLDLLEFAQTTHROHFPDROYVST
LISIKGACPECKYCKPOSSRYKTELEPELMQVBOYLDASRKKNGSPFGCAAM
KNHBRDMPYLEOMVQGVKAMGLACMTGLTDTQORLASAGLDYNNHLDSPER
YCNLTITRETYERLDTLDKVRDAGIKVSGSIVGIGETVYKDRALQLQANLTPPES
VPIKNLVVKGPIADNDVAFDFIRIYAVARIMPPSSVYRLSAGREOMECOTAMC
FMAGANSIFYGCKLITTPNPEEDKDVOLFRKILGINPQOTDVTMGDNEQOKLEBOIEN
ADTDQFYNAAL"
BASE COUNT 230 a 322 c 308 g 181 t
ORIGIN
Query Match 80.8%; Score 20.2; DB 3; Length 1041;
Best Local Similarity 88.0%; Pred. No. 48;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
1 cgatcctcgtcaggtgcagtcagc 25
||||| ||||||| ||||||| |||
Db 105 CGACCTCCGCCAGGTTCCAGTGCAGC 129
RESULT 19
HSY15922
LOCUS HSY15922 160 bp DNA PRI 06-APR-1998
DEFINITION Homo sapiens CTNS gene, exon 1 and 3' flanking intronic region.
ACCESSION Y15922
VERSION Y15922.1 GI:3036848
KEYWORDS CTNS gene; cystinosis.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 160)
AUTHORS Antignac,C.

```

FEATURES	source	location/Qualifiers
JOURNAL	Submitted (15-DEC-1997)	Antignac C., UA23, INSERM, Tour Lavoisier
REFERENCE	Hopital Necker 149 rue de Sevres Paris, 75015, FRANCE	
REFERENCE	2 (bases 1 to 160)	
REFERENCE	Town, M., Jean, G., Cherqui, S., Altard, M., Forestier, L., Whitmore, S.A., Callen, D.F., Gribouval, O., Broyer, M., Bates, G.P., van't Hoff, W. and Antigens, C.	
TITLE	A novel gene encoding an integral membrane protein is mutated in nephropathic cystinosis	
JOURNAL	Nat. Genet. 18 (4), 319-324 (1998)	
FEATURES	98196758	
source	1. 160	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
exon	1. 110	
	/gene="CTNS"	
	/number=1	
	/usedin=Y15924:CTNSmRNA	
gene	1. 110	
	/gene="CTNS"	
	11. >160	
	/number=1	
intron		
BASE COUNT	22 a 55 c 57 g 26 t	
ORIGIN		
Query Match	72.8%; Score 18.2; DB 97; Length 160;	
Best Local Similarity	87.0%; Pred. No. 5.2e+02;	
Matches	20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
Oy	2 gatcctcgtcaggtgcagtcag 24	
	11111111111111111111	
Db	67 GAGCCTCGCAGGTCCGCTCAG 45	
RESULT 20		
LOCUS	CEU10160 525 bp mRNA	09-JUL-1994
DEFINITION	Cervus elaphus interleukin 12 mRNA, partial cds.	
ACCESSION	U10160	
VERSION	U10160.1 GI:508867	
KEYWORDS	red deer.	
SOURCE	Cervus elaphus	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervidae; Cervidae; Cervinae; Cervus.	
REFERENCE	1 (bases 1 to 525)	
REFERENCE	Lockhart, E.A.	
REFERENCE	Cloning and sequencing of cervine IL-12	
REFERENCE	unpublished	
REFERENCE	2 (bases 1 to 525)	
REFERENCE	Lockhart, E.A.	
REFERENCE	Direct Submission	
REFERENCE	Submitted (30-MAY-1994)	
REFERENCE	Euan A. Lockhart, Microbiology, Deer Research Laboratory, University of Otago, Union St., Dunedin, New Zealand	
FEATURES	location/Qualifiers	
source	1. 525	
	/organism="Cervus elaphus"	
	/db_xref="taxon:9606"	
	<1. >525	
	/standard_name="IL-12"	
	/codon_start=2	
	/product="interleukin 12"	
	/protein_id="AA19658.1"	
	/db_xref="GI:508868"	
	/translation="AGQYCHKGGEVLSRLLLHKKEDGIMSTDILKDKQEPKAKSLKCEAKDYSGHGTCMMILTAISTDLKFSYKSGSSDPGVTGCMASISTEKVIVDHRHRYKKTIVCEGSGACPAAEESLPPIEVYVAHVHKLKENTTSFFTRIDIIKDPKPNLQIRPLKNSRQEVYSWE"	
BASE COUNT	154 a 120 c 142 g 109 t	

[illegible]

CEU57752
 LOCUS CEU57752 993 bp mRNA
 DEFINITION Cervus elaphus interleukin-12 p40 subunit mRNA, complete cds.
 ACCESSION U57752
 VERSION U57752.1 GI:1373387
 KEYWORDS
 SOURCE red deer.
 ORGANISM Cervus elaphus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervidae; Cervinae; Cervus.
 REFERENCE 1 (bases 1 to 993)
 Lockhart, E., Slobde, L., and Buchan, G.
 The characterisation of a cervine immunoregulatory cytokine, interleukin 12
 JOURNAL DNA Seq. 10 (2), 139-148 (1999)
 MEDLINE 99304506
 PUBMED 10376218
 REFERENCE 2 (bases 1 to 993)
 Lockhart, E.A. and Buchan, G.S.
 Direct Submission
 JOURNAL Submitted (08-MAY-1996) Euan A. Lockhart, Microbiology, University of Otago, Cumberland St, Dunedin, New Zealand
 FEATURES
 Source
 1. .993
 /organism="Cervus elaphus"
 /db_xref="taxon:9860"
 /cell_type="peripheral blood mononuclear cells"
 primer_bind
 1. .33
 1. .984
 /codon_start=1
 /product="Interleukin-12 p40 subunit"
 /protein_id="AAB02258.1"
 /db_xref="GI:1373388"
 /translation="MHQQLVSNFSLVLTSPVIAIMELKKNVVELDWPDPAGE
 TVVLCDDTPEEDGITWTDOSGEVLGSKTLLTVYKFGAGQYTKHKGGEALSHSL
 LPHKKEDGIWSTDLKDKPEKNFTFLCEKKNYSGRICTWMLSTIDLTFSIISR
 GSSNPGQVTCGAATLSAERVGRDNKVEYSEVCEDSACPAAERLPLEVVDAIHL
 LKENTSSFFRTDITKPPKNIJLPLKNSRQVEVSMEDPTWSTPHSYSLFCIC
 QVQKRNKRKRITFTDKSATVTRKKNASPVQADRYSSNHWMTSVCS"
 complement(961. .993)
 BASE COUNT 269 a 240 c 263 g 221 t
 ORIGIN
 primer_bind
 72.8%; Score 18.2; DB 7; Length 993;
 Best Local Similarity 87.0%; Pred. No. 4.2e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 3 atcctcgtagtgcaggtcagc 25
 ||||| ||||| ||||| |||||
 Db 746 ATTCTCGCAGGTGAGGTGACG 768
 RESULT 24
 LOCUS CTU19834
 DEFINITION CTU19834 1005 bp mRNA
 CERCOCEBUS TORQUATUS INTERLEUKIN-12 ALPHA P40 SUBUNIT (IL-12a)
 mRNA, complete cds.
 ACCESSION U19834
 VERSION U19834.1 GI:644785
 KEYWORDS
 SOURCE red-crowned mangabey.
 ORGANISM Cercocebus torquatus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Cercocebus.
 REFERENCE 1 (bases 1 to 1005)
 Villinger, F., Brar, S.S., Mayne, A., Chikala, N. and Ansari, A.A.
 Comparative sequence analysis of cytokine genes from human and nonhuman primates
 JOURNAL J Immunol. 155 (8), 3946-3954 (1995)
 MEDLINE 96003435

REFERENCE 2 (bases 1 to 1005)
 Villinger, F.J.
 Direct Submission
 JOURNAL Submitted (13-JAN-1995) Francois J. Villinger, Emory University, Pathology, 1327 Clifton Rd, Atlanta, GA 30322, USA
 FEATURES
 source
 1. .1005
 /organism="Cercocebus torquatus"
 /sub_species="atys"
 /db_xref="taxon:9530"
 /clone="Fpn IL-12a"
 /cell_type="peripheral blood mononuclear cells"
 /tissue_type="blood"
 9. .995
 /gene="IL-12a"
 9. .995
 /gene="IL-12a"
 /codon_start=1
 /function="NK cell stimulatory factor"
 /product="interleukin-12 alpha p40 subunit"
 /protein_id="AAA86701.1"
 /db_xref="GI:644786"
 /translation="MCHQQLVSNFSLVLTSPVIAIMELKKNVVELDWPDPAGE
 MVLTCDDTPEEDGITWTDOSGEVLGSKTLLTVYKFGAGQYTKHKGGEALSHSL
 LPHKKEDGIWSTDLKDKPEKNFTFLCEKKNYSGRICTWMLSTIDLTFSIISR
 GSSNPGQVTCGAATLSAERVGRDNKVEYSEVCEDSACPAAERLPLEVVDAIHL
 KYENTSSFFRTDITKPPKNIJLPLKNSRQVEVSMEDPTWSTPHSYSLFCIC
 VQKRNKRKRITFTDKSATVTRKKNASPVQADRYSSNHWMTSVCS"
 BASE COUNT 275 a 252 c 257 g 221 t
 ORIGIN
 Query Match
 72.8%; Score 18.2; DB 91; Length 1005;
 Best Local Similarity 87.0%; Pred. No. 4.2e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 3 atcctcgtagtgcaggtcagc 25
 ||||| ||||| ||||| |||||
 Db 751 ATTCTCGCAGGTGAGGTGACG 773
 RESULT 25
 LOCUS AF180563
 DEFINITION Homo sapiens interleukin 12, p40 mRNA, complete cds.
 ACCESSION AF180563
 VERSION AF180563.1 GI:5923854
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1007)
 Hongyuan, J. and MeiYun, Z.
 Cloning and sequence analysis of IL-12 cDNA from Chinese
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1007)
 Hongyuan, J. and MeiYun, Z.
 Direct Submission
 JOURNAL Submitted (25-AUG-1999) Hepatitis, Institute of Virology, CAPM, 100 Yixingnjie, Xuanwuq, Beijing 100052, China
 FEATURES
 source
 1. .1007
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_type="peripheral mononuclear"
 7. .993
 /note="NKSF/CLMF: IL12p40; 40KD"
 /codon_start=1
 /product="interleukin 12, p40"
 /protein_id="AAB56386.1"
 /db_xref="GI:592385"
 /translation="MCHQQLVSNFSLVLTSPVIAIMELKKNVVELDWPDPAGE


```

variation
      /replace="g"
      878
      /gene="IL-12a"
      /note="Ile to Thr"
      /replace="c"
      895
variation
      /gene="IL-12a"
      /note="Ser to Pro"
      /replace="c"
      263 c
      276 g
      232 t
BASE COUNT      309 a      263 c      276 g      232 t
ORIGIN

Query Match      72.8%; Score 18.2; DB 97; Length 1080;
Best Local Similarity 87.0%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 atccctcagtgagtcagtcagc 25
      |||||||
DB      755 ATTCTGGCAGGTGAGGTCAGC 777

RESULT 31
LOCUS      AX084106      1389 bp      DNA
DEFINITION      Sequence 4 from Patent WO0110912.
ACCESSION      AX084106
VERSION      AX084106.1 GI:13185656
KEYWORDS
SOURCE      synthetic construct.
ORGANISM      artificial construct.
REFERENCE      1 (bases 1 to 1389)
AUTHORS      Gillies,S.D. and Lo,K.M.
TITLE      Multiple cytokine-antibody complexes
JOURNAL      Patent: WO 0110912-A 4 15-FEB-2001;
              Lexigen Pharmaceuticals Corp. (US)
FEATURES
      source
      1.1389
      /organism="synthetic construct"
      /db_xref="taxon:32630"
      /note="human p40-IL-2 fusion protein coding sequence"
BASE COUNT      413 a      325 c      327 g      324 t
ORIGIN

Query Match      72.8%; Score 18.2; DB 10; Length 1389;
Best Local Similarity 87.0%; Pred. No. 4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 atccctcagtgagtcagtcagc 25
      |||||||
DB      743 ATTCTCGCAGGTGAGGTCAGC 765

RESULT 32
LOCUS      AR091391      1399 bp      DNA
DEFINITION      Sequence 3 from patent US 5994104.
ACCESSION      AR091391
VERSION      AR091391.1 GI:10018146
KEYWORDS
SOURCE      unknown.
ORGANISM      unclassified.
REFERENCE      1 (bases 1 to 1399)
AUTHORS      Anderson,R.James, Prentice,H.Grant and MacDonald,I.Duncan.
TITLE      Interleukin-12 fusion protein
JOURNAL      Patent: US 5994104-A 3 30-NOV-1999;
              Location/Qualifiers
              1.1399
              /organism="unknown"
BASE COUNT      390 a      310 c      333 g      366 t

```

```

ORIGIN

Query Match      72.8%; Score 18.2; DB 9; Length 1399;
Best Local Similarity 87.0%; Pred. No. 4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 atccctcagtgagtcagtcagc 25
      |||||||
DB      785 ATTCTCGCAGGTGAGGTCAGC 807

RESULT 33
LOCUS      HUMCLMF40      1399 bp      mRNA
DEFINITION      Human cytotoxic lymphocyte maturation factor 40 kDa subunit mRNA,
              complete cds.
ACCESSION      M65272 M38443 M38444
VERSION      M65272.1 GI:180625
KEYWORDS      cytotoxic lymphocyte maturation factor 40 kDa subunit; Interleukin.
SOURCE      Human, CDNA to mRNA.
ORGANISM      Homo sapiens
REFERENCE      1 (bases 1 to 1399)
AUTHORS      Gubler,U., Chua,A.O., Schoenhaut,D.S., Dwyer,C.M., McComas,W.,
              Moltyka,R., Nabavi,N., Wollitzky,A.G., Quinn,P.M., Familletti,P.C.
              and Gately,M.K.
              Coexpression of two distinct genes is required to generate secreted
              bioactive cytotoxic lymphocyte maturation factor
              Proc. Natl. Acad. Sci. U.S.A. 88, 4143-4147 (1991)
              91239523
FEATURES
      source
      1.1399
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      43.1029
      /codon_start=1
      /product="cytotoxic lymphocyte maturation factor 40 kDa
              subunit"
      /protein_id="AA35695.1"
      /db_xref="GI:180626"
      /translation="MCHQQLVISMFSGLVFLASPLVAIMEKKRYVVELDMYPDARCE
              /translation="MCHQQLVISMFSGLVFLASPLVAIMEKKRYVVELDMYPDARCE
              MVLVLTCDTPEDDGIYTWLIDQSEVLGSGKTLTQVKEFGDAGQYTHKGGEVLSHL
              LHKKEEDGIWSTDLIDQKPKNKTEFLCEAKNYSGEFTCAWLTITSTDLTFVSKSR
              GSSDPQGVGAATLSAERVGDMDKEVEYVECEQEDSACPAAESLPIEVMYDAVKRL
              KYENYSSPFIROIIDKPPKNIOLKPLKNSRQVEVMEYEDPTMSPHSHYSLFLFCVO
              VQGSKREKDRVETDKTSATVYCRKNAISVRAQDRYSSSMENASVPCVS"
BASE COUNT      390 a      310 c      333 g      366 t
ORIGIN

Query Match      72.8%; Score 18.2; DB 97; Length 1399;
Best Local Similarity 87.0%; Pred. No. 4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 atccctcagtgagtcagtcagc 25
      |||||||
DB      785 ATTCTCGCAGGTGAGGTCAGC 807

RESULT 34
LOCUS      AR091395      1560 bp      DNA
DEFINITION      Sequence 11 from patent US 5994104.
ACCESSION      AR091395
VERSION      AR091395.1 GI:10018150
KEYWORDS
SOURCE      unknown.
ORGANISM      unclassified.
REFERENCE      1 (bases 1 to 1560)
AUTHORS      Anderson,R.James, Prentice,H.Grant and MacDonald,I.Duncan.

```



```

AR082684
LOCUS AR082684 2362 bp DNA PAT 01-SEP-2000
DEFINITION Sequence 1 from patent US 5976539.
ACCESSION AR082684
VERSION AR082684.1 GI:10009474
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 2362)
AUTHORS Scott,P. and Trinchieri,G.
TITLE Compositions and methods for use of IL-12 as an adjuvant
JOURNAL Patent: US 5976539-A 1 02-NOV-1999;
FEATURES
Source 1..2362
BASE COUNT 717 a 483 c 542 g 620 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 9; Length 2362;
Best Local Similarity 87.0%; Pred. No. 3.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 atccctgcagtgacgacgac 25
|| ||| ||||| ||||| |||||
Db 775 ATTCTCGCAGGTGCAGGTACG 797

RESULT 39
LOCUS 128325 2362 bp DNA PAT 06-FEB-1997
DEFINITION Sequence 1 from patent US 5571515.
ACCESSION 128325
VERSION 128325.1 GI:1819101
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 2362)
AUTHORS Scott,P. and Trinchieri,G.
TITLE Compositions and methods for use of IL-12 as an adjuvant
JOURNAL Patent: US 5571515-A 1 05-NOV-1996;
FEATURES
Source 1..2362
BASE COUNT 717 a 483 c 542 g 620 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 2362;
Best Local Similarity 87.0%; Pred. No. 3.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 atccctgcagtgacgacgac 25
|| ||| ||||| ||||| |||||
Db 775 ATTCTCGCAGGTGCAGGTACG 797

RESULT 40
LOCUS 189770 2362 bp DNA PAT 10-AUG-1998
DEFINITION Sequence 1 from patent US 5723127.
ACCESSION 189770
VERSION 189770.1 GI:3409710
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 2362)
AUTHORS Scott,P. and Trinchieri,G.
TITLE Compositions and methods for use of IL-12 as an adjuvant
JOURNAL Patent: US 5723127-A 1 03-MAR-1998;

```

```

FEATURES
Source Location/Qualifiers
BASE COUNT 717 a 483 c 542 g 620 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 2362;
Best Local Similarity 87.0%; Pred. No. 3.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 atccctgcagtgacgacgac 25
|| ||| ||||| ||||| |||||
Db 775 ATTCTCGCAGGTGCAGGTACG 797

RESULT 41
LOCUS HSAJ2967 2611 bp mRNA PRI 06-APR-1998
DEFINITION Homo sapiens mRNA for cystinosin.
ACCESSION AJ222967
VERSION AJ222967.1 GI:3036839
KEYWORDS cystinosin; CYSTN gene.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2611)
AUTHORS Antignac,C.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-1997) Antignac C., U423, INSERM, Tour Lavoisier
Hospital Necker 149 rue de Sevres Paris, 75015, FRANCE
2 (bases 1 to 2611)
AUTHORS Town,M., Jean,G., Cherqui,S., Attard,M., Forestier,L.,
Whitmore,S.A., Callen,D.F., Gribouval,O., Broyer,M., Bates,G.P.,
van't Hoff,W. and Antignac,C.
TITLE A novel gene encoding an integral membrane protein is mutated in
nephropathic cystinosis
JOURNAL Nat. Genet. 18 (4), 319-324 (1998)
98196738
FEATURES
Source Location/Qualifiers
1..2611
/organism="Homo sapiens"
/db_xref="taxon:9606"
340..1443
/gene="CTNS"
340..1443
/gene="CTNS"
/codon.start=1
/product="cystinosin"
/protein_id="CA11021.1"
/db_xref="GI:3036840"
/translat="MIRNMLTIFILPLKLVKCESSVLFVPPVYKLENGSSTNVL
TLRPPNATLVITFEITPERSKNITILELDEVVVPGVNSPOVTSQVGLFVYLH
GNSNOTGPRIRLIVIRSSAISTINQVIMIVFVMSISEYOVIMNRKSGVIGISF
DFVALNLTGVATSVENIGILMPYIKEDPLKTPNGVNPVNSDVFESLHAVVFLI
IIVQCCLYERGQGVSVMPAIGLVLALEAFVYTMVAAGVITWLOFLCFSTIKLAV
TIVKYPFOAMNPNFYKSTGMSIGNVLDETFGGSFLQLQMFLOSYNNDDMTLFDPT
KFLGVESIVFDVVFIOHFCLYKRRKPDQLN"
BASE COUNT 525 a 784 c 674 g 628 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 93; Length 2611;
Best Local Similarity 87.0%; Pred. No. 3.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 gatccctgcagtgacgacgac 24
|| ||| ||||| ||||| |||||
Db 67 GAGCCTCGCAGGTGCAGGTACG 45

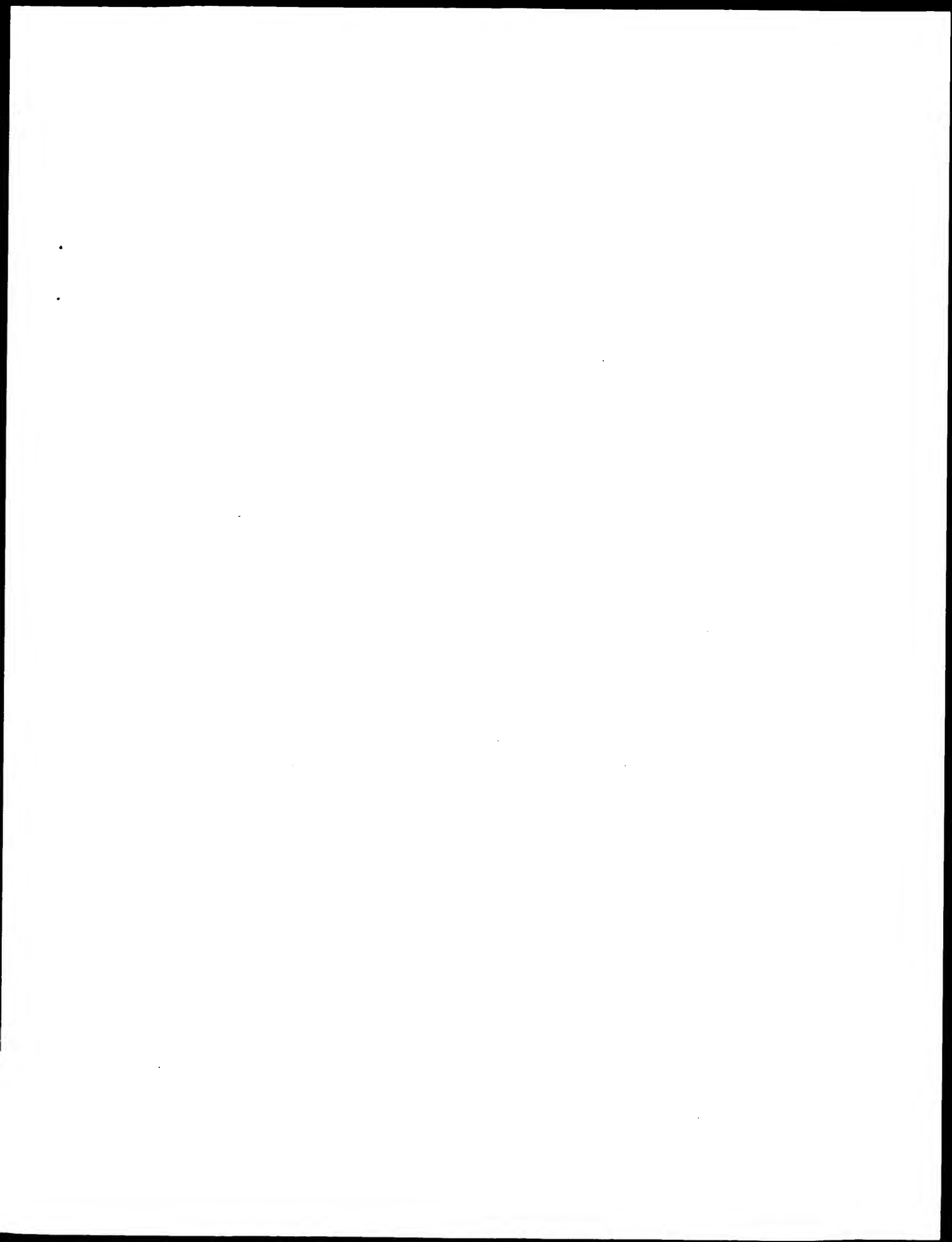
```


*	4250	4677	contig of 428 bp in length
*			gap of unknown length
*	4678	5300	contig of 623 bp in length
*			gap of unknown length
*	5301	5958	contig of 658 bp in length
*			gap of unknown length
*	5959	6598	contig of 640 bp in length
*			gap of unknown length
*	6599	7183	contig of 585 bp in length
*			gap of unknown length
*	7184	7338	contig of 155 bp in length
*			gap of unknown length
*	7339	8020	contig of 682 bp in length
*			gap of unknown length
*	8021	9055	contig of 1035 bp in length
*			gap of unknown length
*	9056	9781	contig of 726 bp in length
*			gap of unknown length
*	9782	10678	contig of 897 bp in length
*			gap of unknown length
*	10679	11487	contig of 809 bp in length
*			gap of unknown length
*	11488	12107	contig of 620 bp in length
*			gap of unknown length
*	12108	12724	contig of 617 bp in length
*			gap of unknown length
*	12725	12906	contig of 182 bp in length
*			gap of unknown length
*	12907	13688	contig of 782 bp in length
*			gap of unknown length
*	13689	14430	contig of 742 bp in length
*			gap of unknown length
*	14431	15569	contig of 1139 bp in length
*			gap of unknown length
*	15570	17161	contig of 1592 bp in length
*			gap of unknown length
*	17162	17335	contig of 114 bp in length
*			gap of unknown length
*	17336	18264	contig of 929 bp in length
*			gap of unknown length
*	18265	19159	contig of 895 bp in length
*			gap of unknown length
*	19160	20056	contig of 897 bp in length
*			gap of unknown length
*	20057	21105	contig of 1049 bp in length
*			gap of unknown length
*	21106	22217	contig of 1112 bp in length
*			gap of unknown length
*	22218	22755	contig of 538 bp in length
*			gap of unknown length
*	22756	23757	contig of 1002 bp in length
*			gap of unknown length
*	23758	24309	contig of 552 bp in length
*			gap of unknown length
*	24310	24642	contig of 333 bp in length
*			gap of unknown length
*	24643	25964	contig of 1322 bp in length
*			gap of unknown length
*	25965	27175	contig of 1211 bp in length
*			gap of unknown length
*	27176	28598	contig of 1423 bp in length
*			gap of unknown length
*	28599	29420	contig of 822 bp in length
*			gap of unknown length
*	29421	30801	contig of 1381 bp in length
*			gap of unknown length
*	30802	32587	contig of 1786 bp in length
*			gap of unknown length
*	32588	33692	contig of 1105 bp in length
FEATURES	Location/Qualifiers		
	1. 33692		
	/organism="Homo sapiens"		
SOURCE	/db_xref="taxon:9606"		

/chromosome="5.5"
 /clone="CTB-140P23"
 BASE COUNT 9394 a 7696 c 7805 g 8776 t 21 others
 ORIGIN
 Query Match 72.8% Score 18.2; DB 61; Length 33692;
 Best Local Similarity 87.0%; Pred. No. 2.7e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 gatactcgtaagtagtcagtcag 24
 ||||| ||||| ||||| |||||
 Db 7341 GACCCTCATACAGCTCAGGTGACG 7363
 RESULT 45
 MTCY71 42729 bp DNA BCT 10-FEB-1999
 DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 141/162.
 ACCESSION Z927771 AL123456
 VERSION Z92777.1 GI:3242259
 KEYWORDS
 SOURCE
 ORGANISM
 Mycobacterium tuberculosis.
 Mycobacterium tuberculosis
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Corynebacterineae; Mycobacteriaceae;
 Mycobacterium; Mycobacterium tuberculosis complex.
 REFERENCE
 AUTHORS
 1 (bases 1 to 4,2729)
 Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
 Harris,D., Gordon,S.V., Eigmeier,K., Gas,S., Barry III,C.E.,
 Tekala,F., Haddock,K., Basham,D., Brown,D., Chillingworth,T.,
 Connor,R., Davies,R., Devlin,K., Fellwell,T., Gentles,S.,
 Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
 Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
 Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
 Squares,S., Sgares,R., Sulston,J.E., Taylor,K., Whitehead,S., and
 Barrell,B.G.
 Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence
 Nature 393 (6685), 537-544 (1998)
 98295987
 Erratum: [[published erratum appears in Nature 1998 Nov
 12;396(6707):1901]]
 2 (bases 1 to 42729)
 Parkhill,J.
 Direct Submission
 Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
 tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
 Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
 Moléculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
 On Jun 20, 1998 this sequence version replaced gi:2143309.
 COMMENT
 Notes:
 Details of M. tuberculosis sequencing at the Sanger Centre are
 available on the World Wide Web.
 (URL, http://www.sanger.ac.uk/projects/M_tuberculosis/) CDS have
 been renumbered from the original cosmid submissions but the old
 gene designations are in brackets after the new gene numbers.
 Gene prediction was based on a Hidden Markov Model of TB genes
 implemented in TBparse (Krogh) supplemented with visual inspection
 of positional base preference in codons, especially where there is
 an increase in the observed/expected third position G + C.
 CAUTION: In some cases we may not have predicted the correct
 initiation codon. Where possible we choose an initiation codon
 (atg, gtg, or ttg) which is preceded by an upstream ribosome
 binding site sequence (optimally 5-13bp before the initiation
 codon). If this cannot be identified we choose the most upstream
 initiation codon.
 Location/Qualifiers
 1..42729
 /organism="Mycobacterium tuberculosis"
 /strain="H37Rv"
 /db_xref="taxon:1773"

Best Local Similarity 87.0%; Pred No. 2.6e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 gatcctcgtcaggtgcaggtcag 24
||||| ||||||| |||
Db 8563 GATCCGCGTCAAGGTGCAGGACTG 8585

Search completed: October 9, 2001, 12:14:40
Job time: 3995 sec



XX Example 2; Column 37-40; 34pp; English.

PS This sequence encodes the E. coli biotin synthetase (BioB). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.

XX Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best local Similarity 100.0%; Pred. No. 0.024;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatcctcgtagtgagtgacgacg 25
 ||||||||||||||||||
 DB 105 cgatcctcgtagtgagtgacgacg 129

RESULT 2

ID AAN91329 standard; DNA; 1084 BP.

AC AAN91329;

DT 15-FEB-1990 (first entry)

DE E.coli Bio B gene.

KW E.coli; Bio B gene; biotin.

OS Escherichia coli.

FH Key Location/Qualifiers

FT CDS 24..1064

PN GB2216530-A.

PD 11-OCT-1989.

PF 17-MAR-1989; 89GB-0006210.

PR 22-MAR-1988; 88GB-0006804.

PR 17-MAR-1989; 89GB-0006210.

PA (UKAG-) UK MIN. AGRIC. FISH.

PI Pearson BW, McKee RA;

DR WPI; 1989-295085/41. P-PSDB P91392

PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes
 PT - derived from E.coli and capable of replication and expression in other
 PT microorganisms, esp. yeast.

XX Table 3; page 33-4; 52pp; English.

CC The gene can be used in a plasmid for expression of enzymes of the biotin
 CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
 CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
 CC Lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.

XX Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best local Similarity 100.0%; Pred. No. 0.025;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatcctcgtagtgagtgacgacg 25
 ||||||||||||||||||
 DB 128 cgatcctcgtagtgagtgacgacg 152

RESULT 3

ID AAN60496 standard; DNA; 1121 BP.

AC AAN60496;

DT 17-OCT-1991 (first entry)

DE Sequence encoding biotin synthesising enzyme.

KW Biotin synthetic enzyme; E.coli; desbiobiotin; ds.

FH Key Location/Qualifiers

FT CDS 42..1082

PN JP61149091-A.

PD 07-JUL-1986.

PF 24-DEC-1984; 84JP-0272605.

PR 24-DEC-1984; 84JP-0272605.

PA (NIPS) NIPPON SODA KK.

DR WPI; 1986-216622/33.

DR P-PSDB; AAP60536.

PT Double stranded DNA encoding biotin synthesising enzyme -
 PT comprises transformed mutant E.coli strain contg. cyclic doubled
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.

PS Disclosure; Page 534; 23pp; Japanese.

CC The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing desbiobiotin.

XX Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best local Similarity 100.0%; Pred. No. 0.025;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatcctcgtagtgagtgacgacg 25
 ||||||||||||||||||
 DB 146 cgatcctcgtagtgagtgacgacg 170

RESULT 4

ID AAO62386 standard; DNA; 5872 BP.

AC AAO62386;

DT 16-NOV-1994 (first entry)

DE Biotin-biosynthesis genes contg. plasmid pB030A-15/9.

KW Biotin; expression; enterobacteria; vitamin H; synthesis;
 KW plasmid; pB030A-15/9; bioB; bioC; bioD; bioA;
 KW promoter plac; biotin synthase; KAPA synthase;
 KW 8-amino-7-oxononanoate synthase; pimeoyl-coA; DTB synthase;


```

KW  dethiobiotin synthase; DAPA synthase;
KM  S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
XX  seborrhoea; dermatitis; ds.
XX  Escherichia coli DSM498.
FH  key
FT  Location/Qualifiers
FT  1..96
FT  /tag= a
FT  /function= "promoter ptac"
FT  /evidence= EXPERIMENTAL
FT  23..28
FT  /tag= b
FT  /standard_name= "promoter ptac"
FT  45..50
FT  /tag= c
FT  /evidence= EXPERIMENTAL
FT  /standard_name= "promoter ptac"
FT  105..109
FT  /tag= d
FT  /evidence= EXPERIMENTAL
FT  /standard_name= "biob RBS no. 9"
FT  117..1157
FT  /tag= e
FT  /product= "biotin synthase"
FT  /evidence= EXPERIMENTAL
FT  /gene= "biob"
FT  /number= 1
FT  1141..1146
FT  /tag= f
FT  /standard_name= "biof RBS"
FT  1154..2311
FT  /tag= g
FT  /EC_number= 2.3.1.47
FT  /product= "KAPA synthase"
FT  /evidence= EXPERIMENTAL
FT  /gene= "biof"
FT  /number= 2
FT  /standard_name= "8-amino-7-oxononanoate synthase"
FT  2284..2288
FT  /tag= h
FT  /standard_name= "bioc RBS"
FT  2295..3050
FT  /tag= i
FT  /function= "involved in pimeloyl-CoA synthesis"
FT  /product= "protein"
FT  /gene= "bioc"
FT  /number= 3
FT  3030..3033
FT  /tag= j
FT  /standard_name= "bioc RBS"
FT  3043..3753
FT  /tag= k
FT  /EC_number= 6.3.3.3
FT  /product= "DTB synthase"
FT  /evidence= EXPERIMENTAL
FT  /gene= "bioid15"
FT  /number= 4
FT  /standard_name= "dethiobiotin synthase"
FT  3712..3750
FT  /tag= l
FT  /note= "bioid15 substitution"
FT  3742..3746
FT  /tag= m
FT  /standard_name= "bioc RBS"
FT  3750..5039
FT  /tag= n
FT  /EC_number= 2.6.1.62
FT  /product= "DAPA synthase"
FT  /evidence= EXPERIMENTAL
FT  /gene= "bioc"
FT  /number= 5
FT  /standard_name= "S-adenosyl-L-methionine: 8-amino-

```

```

FT  RBS
FT  5088..5093
FT  /tag= o
FT  /standard_name= "ORF1 RBS"
FT  5098..5574
FT  /tag= p
FT  /function= "unknown, involved in biotin synthesis"
FT  /product= "protein"
FT  /evidence= EXPERIMENTAL
FT  /gene= "ORF1"
FT  /number= 6
FT  5583..5644
FT  /tag= q
FT  /standard_name= "rho-independent transcriptional
FT  terminator"
FT  5583..5605
FT  /tag= r
FT  stem_loop
FT  5583..5605
FT  /tag= r
XX  WO9408023-A.
XX  14-APR-1994.
XX  01-OCT-1993; 93WO-EP02688.
XX  02-OCT-1992; 92CH-0003124.
XX  15-JUL-1993; 93CH-0002134.
XX  (LONZ ) LONZA AG.
XX  PA
XX  PI
XX  Birch O, Brass J, Fuhrmann M, Shaw N;
XX  WPI: 1994-135587/16.
XX  DR  P-PSDB: AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
XX  PT  Biotechnological biotin prodn. using enterobacterial biotin-gene
XX  - providing vitamin H in high yield
XX  PS  Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
XX  CC  The sequence is derived from plasmid pB030A-15/9 contg. the
XX  CC  biob, biof, bioc, bioid and bioc genes responsible for biosynthesis
XX  CC  contg. these DNA fragments or plasmids may be used in the prodn.
XX  CC  of biotin. Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
XX  CC  loss of appetite and tiredness.
XX  SQ  Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;
XX  Query Match 100.0%; Score 25; DB 15; Length 5872;
XX  Best Local Similarity 100.0%; Pred. No. 0.029;
XX  Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX  QY  1 cgaactcgtcagtgagtcagc 25
XX  |||||||
XX  Db 221 cgaactcgtcagtgagtcagc 245
XX  RESULT 5
XX  ID AA004701
XX  AA004701 standard; DNA; 906 BP.
XX  AC AA004701;
XX  AC
XX  DT 11-OCT-1990 (first entry)
XX  DE
XX  KW Sequence encoding natural killer cell stimulatory factor (NKSF).
XX  KW Natural killer cell stimulatory factor: NKSF; cancer;
XX  KW leukaemia; anaemia; radiation; bone marrow transplantation; ds.
XX  OS Homo sapiens.
XX  XX

```

```

PN W09005147-A.
XX
XX 17-MAY-1990.
XX
XX 09-NOV-1989; 89WO-0005027.
XX
XX 10-NOV-1988; 88US-0269945.
XX
XX (GENE-) GENETICS INST INC.
XX (WIST-) WISTAR INST.
XX
XX Tinchieri G, Perussia B, Kobayashi M, Clark SC, Wong GC;
XX WPI; 1990-178800/23.
XX P-PSDB; AAR05171.
XX
XX New natural killer cell stimulatory factor -
XX useful for treating cancer and eg. leukopenia, bacterial and
XX viral infections, anaemia and B- or T-cell deficiencies.
XX
XX Claim 16; Page 60; 69pp; English.
XX
XX Natural killer cell stimulatory factor (NKSF) is useful in
XX treating cancer and other disease states which respond to NK cell
XX activity. Such states may arise from immune or haematopoietic-cell
XX deficiency following a bone marrow transplantation, exposure to
XX radiation and disease.
XX
XX Sequence 906 BP; 244 A; 220 C; 238 G; 204 T; 0 other;
XX
XX
XX Query Match 72.8%; Score 18.2; DB 11; Length 906;
XX Best Local Similarity 87.0%; Pred. No. 28;
XX Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 3 atctctgcaggtgcagtcagc 25
XX 11111111111111111111
XX Db 775 attctcgacagtgagagtcagc 797
XX
XX RESULT 6
XX AAQ12149
XX ID AAQ12149 standard; DNA; 979 BP.
XX
XX AC AAQ12149;
XX
XX DT 10-SEP-1991 (first entry)
XX
XX DE Cytotoxic lymphocyte maturation factor (40 kD subunit) gene.
XX
XX KW Cytotoxic lymphocyte maturation factor; CLMF; probe; antitumour;
XX monoclonal antibodies; T cell; ss.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX FT CDS 41..979
XX FT /tag= a
XX FT /product= CLMF
XX FT 41..106
XX FT /tag= b
XX FT /note= "hydrophobic signal peptide"
XX FT 107..979
XX FT /tag= c
XX
XX EP433827-A.
XX
XX PD 26-JUN-1991.
XX
XX PF 09-DEC-1990; 90EP-0123670.
XX
XX PR 27-AUG-1990; 90US-0572284.
XX 22-DEC-1989; 89US-0455708.
XX

```

```

PR 09-MAY-1990; 90US-0520935.
XX
XX (HOFF ) HOFFMANN-LA ROCHE AG.
XX
XX Chizzonite RA, Gately MK, Gubler UA, Hulmes JD, Pan YCE;
XX Podlaski FJ, Stern AS;
XX
XX WPI; 1991-186715/26.
XX P-PSDB; AAR12493.
XX
XX New cytotoxic lymphocyte maturation factor - used as antitumour
XX agent and monoclonal antibodies used in transplantation to block
XX proliferation and activation of cytotoxic T cells
XX
XX Disclosure; Fig 25 (a-c); 90pp; English.
XX
XX The natural CLMF protein is a 75 kD heterodimer comprising two
XX polypeptide subunits, a 40 kD and a 35 kD subunit (AAQ12150).
XX The subunits are bonded together via one or more disulfide bonds.
XX The CLMF can interact with IL-2 to synergistically induce the
XX cytolytic activity of lymphokine Activated Killer (LAK) cells. It is
XX also capable of stimulating T-cell growth. It can be used for the
XX prepn. of LAK- and T-cell activator and antitumour compsns. and for
XX stimulating Natural Killer (NK) cells.
XX
XX Sequence 979 BP; 261 A; 245 C; 257 G; 216 T; 0 other;
XX
XX
XX Query Match 72.8%; Score 18.2; DB 12; Length 979;
XX Best Local Similarity 87.0%; Pred. No. 28;
XX Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX QY 3 atctctgcaggtgcagtcagc 25
XX 11111111111111111111
XX Db 783 attctcgacagtgagagtcagc 805
XX
XX RESULT 7
XX AAT50759
XX ID AAT50759 standard; cDNA; 984 BP.
XX
XX AC AAT50759;
XX
XX DT 24-SEP-1997 (first entry)
XX
XX DE Sequence of ovine IL-12 40 kD subunit cDNA.
XX
XX KW Cytokine; ovine; sheep; interleukin-5; interleukin-12; IL-5; IL-12;
XX livestock; cow; stress; transport; vaccine adjuvant; veterinary;
XX cancer; immunosuppression; allergy; reproductive system; growth;
XX early maturity; antibody; diagnosis; immunopotentiator;
XX early haematopoietic progenitor cell; cytotoxic cell; thymocyte;
XX secretion; IgM; IgA; bacterial endotoxin; gamma-interferon; ss.
XX
XX OS Ovis ammon aries.
XX
XX FH Ovis ammon aries.
XX FT WO9700321-A1.
XX
XX PD 03-JAN-1997.
XX
XX PF 14-JUN-1996; 96WO-AU00360.
XX
XX PR 27-OCT-1995; 95AU-0006244.
XX PR 14-JUN-1995; 95AU-0003502.
XX
XX (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX
XX Seow H, Wood P;
XX
XX WPI; 1997-077528/07.
XX P-PSDB; AAM08483.
XX
XX Nucleic acid encoding ovine interleukin-5 or -12 - used as vaccine
XX

```

PT adjuvants and to treat or prevent microbial infections in livestock
 XX
 PS Claim 10; Page 49-51; 78pp; English.
 CC The sequences given in AAT50757-59 encode ovine interleukin-12 (IL-12)
 CC 35 kD subunit (partial and full length sequence) and the 40 kD subunit.
 CC Ovine IL-5 or IL-12 are used to treat and/or prevent infections in
 CC livestock (esp. cows and sheep), particularly where the animals are
 CC stressed, e.g. during transport. IL-5 and IL-12 can also be used as
 CC adjuvants in vaccines for veterinary use (partic. weakly immunogenic
 CC subunit or synthetic peptide vaccines). They may also be used to
 CC treat cancer, immunosuppression and allergy, to enhance/suppress the
 CC reproductive system and to promote growth or early maturity. Optionally
 CC interleukin can be delivered from constructs or delivery cells and
 CC antibodies are useful in enzyme immunoassays for rapid diagnosis of
 CC infection. The interleukins are immunopotentiators, especially IL-5
 CC promotes growth of early haematopoietic progenitor cells and generation
 CC of cytotoxic cells from thymocytes, also it stimulates production and
 CC secretion of IgM and IgA (in synergism with bacterial endotoxin).
 CC IL-12 induces production of gamma-interferon by, and proliferation
 CC of, T and NK cells and increases the (non-)specific cytolytic
 CC lymphocyte response. The genetic constructs can also be used for
 CC in vitro production of IL-5 or -12.
 XX
 SQ Sequence 984 BP; 273 A; 243 C; 253 G; 215 T; 0 other;

Query Match 72.8%; Score 18.2; DB 18; Length 984;
 Best Local Similarity 87.0%; Pred. No. 28;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 atctcgtcaggtgcagtcagc 25
 || |||| |||| |||| ||||
 Db 746 atctcgcagcagtgagtcagc 768

RESULT 8
 ID AAO97484 standard; cDNA; 987 BP.
 AC AAO97484;
 XX
 DT 22-DEC-1995 (first entry)
 DE Human interleukin-12 40 kD subunit.
 DE
 XX
 KW Interleukin-12; IL-12; cytokine; gamma-interferon; T cell;
 KW natural killer cell; graft vs. host disease; ss.
 KW
 XX
 OS Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS 1..987
 FT /*tag= a
 FT
 XX
 PN WO9519786-A.
 XX
 PD 27-JUL-1995.
 PD
 XX
 PF 20-JAN-1995; 95WO-US00879.
 PF
 XX
 PR 21-JAN-1994; 94US-0186529.
 PR
 XX
 PA (GENE-) GEN HOSPITAL CORP.
 PA (GENE-) GENETICS INST INC.
 XX
 PI Sykes M, Wolf SF;
 PI
 XX
 DR WPI: 1995-269273/35.
 DR P-PSDB: AAR79186.
 DR
 PT Control of graft versus host disease using interleukin 12 - partic.
 PT admin. at the time of bone marrow transplant

XX Disclosure: Page 21-22; 32pp; English.
 PS
 XX AAO97484 and AAO97485 are from human lymphoblast cell line RPMI 8866.
 CC IL-12 is a heterodimeric cytokine which was originally identified as a
 CC factor which induces gamma-interferon from T cells and natural
 CC killer cells as set forth in PCT/US91/06332, published April 2 1992,
 CC and incorporated by reference. EP 433827, published June 26 1991,
 CC disclosed IL-12 as a cytotoxic lymphocyte maturation factor. In the
 CC patent, IL-12 is defined as a heterodimeric glycoprotein comprised
 CC of two covalently linked subunits having mol. wts of about 40 kD and
 CC 35 kD. The two subunits may be disulfide-bonded. IL-12 is
 CC administered to prevent or ameliorate GVHD. Dosage is 1 ng-0.1 mg
 CC (esp. 1.0 ng-1 mcg)/kg daily for 3 days, the first dose given
 CC 1 hr. after transplant. IL-12 can be used as the native heterodimer
 CC or as subunits.
 XX
 SQ Sequence 987 BP; 264 A; 243 C; 259 G; 221 T; 0 other;

Query Match 72.8%; Score 18.2; DB 16; Length 987;
 Best Local Similarity 87.0%; Pred. No. 28;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 atctcgtcaggtgcagtcagc 25
 || |||| |||| |||| ||||
 Db 743 atctcgcagcagtgagtcagc 765

RESULT 9
 ID AAO7273 standard; cDNA; 987 BP.
 AC AAO7273;
 XX
 DT 25-SEP-1998 (first entry)
 DE Human Interleukin-12 p40 gene.
 DE
 XX
 KW IL-12 subunit; expression construct; treatment: asthma; microbial
 KW infection; viral infection; cancer; Human; Interleukin; ss.
 KW
 XX
 OS Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS 1..987
 FT /*tag= a
 FT /product= "IL-12 p40"
 FT
 XX
 PN WO9817689-A2.
 XX
 PD 30-APR-1998.
 PD
 XX
 PF 10-OCT-1997; 97WO-US18779.
 PF
 XX
 PR 18-OCT-1996; 96US-0028676.
 PR
 XX
 PA (GENE-) GENEMEDICINE INC.
 PA
 XX
 PI Deshpande D, Freimark B, Nordstrom J;
 PI
 XX
 DR WPI: 1998-261428/23.
 DR P-PSDB: AAW51311.
 DR
 PT Constructs for expression of interleukin-12 sub-units - are used
 PT for delivery of IL-12 sub-units for treating e.g. asthma, microbial
 PT or viral infections and certain cancers
 XX
 PS Claim 1; Page 34; 80pp; English.
 XX
 CC The IL-12 subunits AAO7273 and AAO7274 can be expressed using an
 CC appropriate expression construct engineered into a plasmid. The
 CC plasmids can provide for efficient expression of IL-12 subunits. The


```

XX OS Mouse; immunodeficient; pathogen; tumour; lymphocyte; antigen;
KW immunomodulator; vector; vaccine; cancer; HIV; Leishmania;
KW Mycobacterium; listeria; plasmidium; retrovirus; evaluation;
KW human immunodeficiency virus; ds.
XX OS Homo sapiens.
XX PN WO9844788-A2.
XX PD 15-OCT-1998.
XX PF 09-APR-1998; 98WO-US06944.
XX PR 09-DEC-1997; 97US-0069163.
XX PR 09-APR-1997; 97US-0838702.
XX PR 01-MAY-1997; 97US-0848760.
XX PA (CHAN/) CHANG L.
XX PI Chang L;
XX DR WPI: 1999-024005/02.
XX PT Use of immunodeficient mice comprising human cells - particularly
XX SC/D/beige mice comprising human immune cells for evaluating
XX PT vaccines against cancers or human pathogens, e.g. HIV
XX PS Example 10(c)(1); Page 110; 154pp; English.
XX CC Immunodeficient mice comprising human cells can be used for
XX CC exposure to human pathogens and/or their components or human
XX CC tumour cells and human peripheral blood lymphocytes. Also claimed
XX CC is a vaccine comprising a cell modified to express an antigen and
XX CC an immune-modulating protein, this is preferably an expression
XX CC vector comprising a polynucleotide sequence that encodes the
XX CC antigen and immune-modulating protein. Such vectors can be used
XX CC to treat a subject having a tumour by transferring the expression
XX CC vector into the tumour so that the antigen and the immune-modulator
XX CC are expressed by at least the tumour. The methods can be used for
XX CC producing and evaluating vaccines including cancer vaccines and
XX CC vaccines directed against human pathogens, e.g. HIV, Leishmania,
XX CC Mycobacterium, listeria or plasmidium. Two primers (AAV83197,
XX CC AAV83198) were used to amplify the IL-12B for insertion into
XX CC expression vectors of the invention.
XX SQ Sequence 987 BP; 264 A; 245 C; 257 G; 221 T; 0 other;

Query Match 72.8%; Score 18.2; DB 20; Length 987;
Best Local Similarity 87.0%; Pred. No. 28;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 atctctgcaggtgcagtcagc 25
   || ||| ||||| ||||| |||
DB 743 attctgcagcagtgtagtcagc 765

RESULT 13
ID AA067879
AC AA067879;
XX 23-MAR-1995 (first entry)
DE EPV 42 KD promoter/Human IL-12 p40 subunit expression cassette.
XX
KW Polymerase chain reaction; primer; amplify; NYVAC; ALVAC; recombinant;
KW murine; interleukin-2; IL-2; pRW825; pmut-1; pBS-SK; PMM151; TK vector;
KW plasmid; vaccinia; H6 promoter; amplify; primer; antigenic response;
KW polymerase chain reaction; poxvirus; PSD542; immunological response;
KW pathogen; human; interferon; IFN; ss.

```

```

XX OS Synthetic.
XX EH Key
XX FT promoter
XX FT CDS
XX FT
XX FT
XX FT
XX PN WO9416716-A.
XX PD 04-AUG-1994.
XX PF 21-JAN-1994; 94WO-US00888.
XX PR 21-JAN-1993; 93US-0007115.
XX PR 19-JAN-1994; 94US-0184009.
XX PA (VIRO-) VIROGENETICS CORP.
XX PI Cox WI, Paoletti E, Tartaglia J;
XX DR WPI: 1994-263767/32.
XX PT Attenuated recombinant virus used for cancer therapy - comprises
XX PT DNA encoding cytokine and/or tumour associated antigen
XX PS Example 27; Fig 33; 232pp; English.
XX CC The sequences given in AA067879-80 are expression cassettes which
XX CC contain the p40 and p35 subunits of human interleukin-12 (IL-12) under
XX CC the control of the entomopox virus (EPV) promoter and the vaccinia E3L
XX CC promoter, respectively. These sequences were used in the construction
XX CC of an ALVAC-based recombinant virus containing the genes for the two
XX CC subunits of human IL-12. The p40 subunit was amplified using primers
XX CC JP202 and JP189 and placed under the control of the entomopox virus
XX CC (EPV) 42 KD promoter in plasmid PBSH112p40II. The p35 subunit was
XX CC amplified using primers JP186 and JP201 and placed under the control
XX CC of the vaccinia E3L promoter using primers JP218 and JP220. The
XX CC expression cassette was isolated and ligated with pBSK+ generating
XX CC plasmid PBSH112p35II. The p35 and p40 expression cassettes were
XX CC excised and assembled in pBSK+ to give PBSH112 in which the two
XX CC expression cassettes are orientated head-to-head relative to each other.
XX CC The combination cassette containing the poxvirus promoted genes for both
XX CC IL-12 subunits was excised from PBSH112 and cloned in to ALVAC C6
XX CC vector plasmid pC6L to give pC6H112. Recombination between this donor
XX CC plasmid and ALVAC rescuing virus produced recombinant virus ALVAC+IL12,
XX CC which contains both the human IL-12 genes in the C6 locus. The resulting
XX CC virus may be used in a composition for inducing an antigenic or
XX CC immunological response, ie. for immunisation against pathogens.
XX SQ Sequence 1018 BP; 283 A; 245 C; 260 G; 230 T; 0 other;

Query Match 72.8%; Score 18.2; DB 15; Length 1018;
Best Local Similarity 87.0%; Pred. No. 28;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 atctctgcaggtgcagtcagc 25
   || ||| ||||| ||||| |||
DB 774 attctgcagcagtgtagtcagc 796

RESULT 14
ID AA208510
AC AA208510;
XX 19-OCT-1999 (first entry)
XX

```

DE BPV 42 kDa/human IL-12 p40 expression cassette.
 XX
 KW Attenuated recombinant virus; cytokine; tumour associated antigen;
 KW NYVAC recombinant virus; ALVAC recombinant virus; gene therapy; rabies;
 KW cancer; tumour necrosis factor; nuclear phosphoprotein; p53; IL-2; GM-CSF;
 KW Interleukin; Interferon; IFN-gamma; IL-4; melanoma associated antigen;
 KW carcinoembryonic antigen; immunisation; antigenic; poxvirus; Influenza;
 KW Immunological response; immunotherapy; vaccine; Newcastle Disease; ss.
 XX
 OS Synthetic.
 OS Entomopox virus.
 OS Homo sapiens.
 XX
 PN US5942235-A.
 XX
 PD 24-AUG-1999.
 XX
 PF 02-JUN-1995; 95US-0458356.
 XX
 PR 02-JUN-1995; 95US-0458356.
 PR 24-DEC-1981; 81US-0334456.
 PR 08-DEC-1982; 82US-0446824.
 PR 19-JUN-1984; 84US-0622135.
 PR 27-AUG-1987; 87US-0090209.
 PR 28-AUG-1987; 87US-0090711.
 PR 20-OCT-1987; 87US-0110335.
 PR 25-APR-1988; 88US-0186054.
 PR 23-AUG-1988; 88US-0234390.
 PR 08-MAR-1989; 88US-0320471.
 PR 14-FEB-1990; 90US-0478179.
 PR 14-JUN-1990; 90US-0537882.
 PR 14-JUN-1990; 90US-0537890.
 PR 07-JAN-1991; 91US-0638080.
 PR 07-MAR-1991; 91US-0666056.
 PR 11-JUN-1991; 91US-0713967.
 PR 16-DEC-1991; 91US-0805567.
 PR 03-MAR-1992; 92US-0847977.
 PR 06-MAR-1992; 92US-0847951.
 PR 22-JUL-1992; 92US-0918278.
 PR 20-JAN-1993; 93US-0007115.
 PR 19-JAN-1994; 94US-0184009.
 PR 14-APR-1994; 94US-0228926.
 PR 13-SEP-1994; 94US-0306259.
 PA
 PA (HEAL-) HEALTH RES INC.
 XX
 PI Paolletti E.
 XX
 DR WPI: 1999-493494/41.
 XX
 PT Recombinant poxviruses comprising exogenous DNA encoding antigenic
 PT determinants useful in immunotherapy to immunize against cancers and
 PT other diseases such as influenza, Newcastle Disease and rabies
 XX
 PS Example 27; Fig 33; 163pp; English.
 XX
 CC The present invention describes a recombinant poxvirus (I), comprising
 CC exogenous DNA encoding an antigenic determinant of a pathogen which is
 CC then expressed in vivo in infected host cells after administration to a
 CC patient and therefore induces an immunological response. (I) may be used
 CC to vaccinate patients against a wide range of diseases and disorders
 CC depending on the type of antigen encoded by the exogenous DNA. (I) may
 CC be used to vaccinate against diseases such as rabies, influenza and
 CC Newcastle Disease. It is particularly useful for immunising against
 CC lymphocytes and tumour cells for use in cell-based immunotherapeutic
 CC modalities for cancer. (I) also have enhanced safety compared to
 CC unattenuated viruses (attenuation reduces the virulence of the viruses)
 CC and known recombinant poxvirus vaccines. This increased level of safety
 CC reduces the possibility of a 'runaway' infection in the host and reduces
 CC the chance of transmission from vaccinated to unvaccinated individuals
 CC and contamination of the environment. The present sequence represents a

CC BPV 42 kDa/human IL-12 p40 expression cassette used in the
 CC exemplification of the present invention.
 XX
 SQ Sequence 1018 BP; 283 A; 245 C; 260 G; 230 T; 0 other;
 XX
 Query Match 72.8%; Score 18.2; DB 21; Length 1018;
 Best Local Similarity 87.0%; Pred. No. 28;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 XX
 QY 3 atctcgtcaggtgaggtcagc 25
 DB 774 atctcgtcaggtgaggtcagc 796
 XX
 RESULT 15
 AAA70709
 ID AAA70709 standard; DNA; 1019 BP.
 XX
 AC AAA70709;
 XX
 DT 15-DEC-2000 (first entry)
 XX
 DE Bovine interleukin-12 p40 subunit gene.
 XX
 KW Antifective; bovine; Interleukin 12; IL12; heterodimer; infection; ss.
 XX
 OS Bos taurus.
 XX
 FH Key location/Qualifiers
 FT CDS 12..995
 FT /*tag= a
 FT /product= "IL-12 p40 subunit"
 XX
 PN JP2000157274-A.
 XX
 PD 13-JUN-2000.
 XX
 PE 20-NOV-1998; 98JP-0331052.
 XX
 PR 20-NOV-1998; 98JP-0331052.
 XX
 PA (DAUC) DAIICHI PHARM CO LTD.
 XX
 DR WPI: 2000-468201/41.
 DR P-PSDB; AAB15391.
 XX
 PT Preparation of bovine interleukin 12 comprising using a vector
 PT containing the gene encoding it, useful for the prevention and
 PT treatment of chronic and opportunistic infections -
 XX
 PS Example 1; Page 8-9; 11pp; Japanese.
 XX
 CC The invention relates to a method for the preparation of bovine
 CC interleukin 12 (BoIL12). BoIL12 consists of a heterodimer of p35
 CC and p40 subunits. BoIL12 is produced by introducing into a host cell,
 CC recombinant vectors containing genes encoding the p35 and p40 subunits.
 CC This sequence represents the coding region for the bovine IL12 p40
 CC subunit. Interleukin 12 is used for the prevention and treatment of
 CC chronic infections and opportunistic infections.
 XX
 SQ Sequence 1019 BP; 273 A; 248 C; 271 G; 227 T; 0 other;
 XX
 Query Match 72.8%; Score 18.2; DB 21; Length 1019;
 Best Local Similarity 87.0%; Pred. No. 28;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 3 atctcgtcaggtgaggtcagc 25
 DB 757 atctcgtcaggtgaggtcagc 779

```

RESULT 16
AAF76880
ID AAF76880 standard; DNA: 1389 BP.
XX
AC AAF76880;
XX
DT 15-MAY-2001 (first entry)
XX
DE Human p40-IL-2 coding sequence.
XX
KW Human p40-IL-2; interleukin-2; IL-2; antiviral; antibacterial;
XX antifungal; cytostatic; antiparasitic; vaccine; gene therapy;
XX fusion protein; protein complex; cytokine; immunoglobulin; cancer;
XX infection; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WC200110912-A1.
XX
PD 15-FEB-2001.
XX
PF 09-AUG-2000; 2000WO-US21715.
XX
PR 09-AUG-1999; 99US-0147924.
XX
PA (LEXI-) LEXIGEN PHARM CORP.
XX
PI Gillies SD, Lo KM;
XX
DR WPI: 2001-191523/19.
XX
PT Novel multifunctional fusion protein or protein complexes useful for
XX treating cancer and viral infections, comprise two different cytokine
XX molecules and a targeting group -
XX
PS Example 1; Page 75-76; 90pp; English.
XX
CC The present sequence is given in a specification relating to
XX multifunctional fusion proteins and protein complexes comprising at
XX least two different cytokine molecules and a targeting group such as an
XX immunoglobulin. The fusion proteins and protein complexes are useful for
XX treating cancers, and bacterial, viral, fungal and parasitic infections.
XX They are also useful in gene therapy and as vaccines to prevent or treat
XX various diseases.
XX
SQ Sequence 1389 BP; 413 A; 325 C; 327 G; 324 T; 0 other;

Query Match          72.8%; Score 18.2; DB 22; Length 1389;
Best Local Similarity 87.0%; Pred. No. 29;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 atctctgcagtgatgcagtcagc 25
   || |||| |||| |||| |||| ||
DB 743 attctgcaggtgagtcagc 765

RESULT 17
AAV00401
ID AAV00401 standard; DNA: 1397 BP.
XX
AC AAV00401;
XX
DT 18-MAR-1998 (first entry)
XX
DE 40 kDa subunit of human cytotoxic lymphocyte maturation factor DNA.
XX
KW Cytotoxic lymphocyte maturation factor; CLMF; 40 kDa subunit;
XX interleukin-2; T-cell growth; antibody; lymphokine activated killer cell;
XX CLMF bioactivity inhibition; affinity ligand; transplantation;
XX cytotoxic T cell proliferation; ss.
XX

```

```

OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 41..1027
XX FT /*tag= a
XX
PN EP790309-A1.
XX
PD 20-AUG-1997.
XX
PF 09-DEC-1990; 90EP-0123670.
XX
PR 27-AUG-1990; 90US-0572284.
XX 22-DEC-1989; 89US-0455708.
XX 09-MAY-1990; 90US-0520935.
XX
PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
XX
PI Chizzonite RA, Gately MK, Gubler UA, Hulmes J, Pan YE;
XX Podlanski FJ, Stern AS;
XX
DR WPI: 1997-404734/38.
XX
P-PSDB; AAW36941.
XX
PT Sub-unit of cytotoxic lymphocyte maturation factor protein - useful
XX for producing CLMF protein and as an antitumour agent
XX
PS Disclosure: Fig 25; 83pp; English.
XX
CC The present DNA sequence encodes a novel 40 kDa subunit of cytotoxic
XX lymphocyte maturation factor (CLMF) protein. The natural CLMF protein is
XX a 75 kDa heterodimer, where the 2 subunits (see AAV00402 for the 35 kDa
XX subunit) are bonded together via one or more disulphide bonds. CLMF
XX protein is produced by a human B lymphoblastoid cell line. CLMF
XX synergistically induces, in the presence of interleukin-2, the cytolytic
XX activity of lymphokine activated killer cells. CLMF is also capable of
XX stimulating T-cell growth. The subunits may be used for preparing
XX the CLMF protein and for preparing medicaments for antitumour therapy.
XX
SQ Sequence 1397 BP; 390 A; 307 C; 333 G; 367 T; 0 other;

Query Match          72.8%; Score 18.2; DB 18; Length 1397;
Best Local Similarity 87.0%; Pred. No. 29;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 atctctgcagtgatgcagtcagc 25
   || |||| |||| |||| |||| ||
DB 783 attctgcaggtgagtcagtcagc 805

RESULT 18
AAT77849
ID AAT77849 standard; DNA: 1397 BP.
XX
AC AAT77849;
XX
DT 18-MAR-1998 (first entry)
XX
DE 40 kDa subunit of human cytotoxic lymphocyte maturation factor DNA.
XX
KW Cytotoxic lymphocyte maturation factor; CLMF; 40 kDa subunit;
XX interleukin-2; T-cell growth; antibody; lymphokine activated killer cell;
XX CLMF bioactivity inhibition; affinity ligand; transplantation;
XX cytotoxic T cell proliferation; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 41..1027
XX FT /*tag= a
XX
PN EP790255-A2.
XX

```

```

XX 20-AUG-1997.
PD 09-DEC-1990; 90EP-0123670.
XX
XX 27-AUG-1990; 90US-0572284.
PR 22-DEC-1989; 89US-0455708.
PR 09-MAY-1990; 90US-0520935.
XX
PA (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI Chizzoneite RA, Gately MK, Gubler UA, Hulmes JD;
PI Pan YE, Podlanski FJ, Stern AS;
XX
DR WPI: 1997-404698/38.
DR P-PSDB; AAM24235.
XX
PT Antibodies to cytotoxic lymphocyte maturation factor protein -
PT useful as affinity ligands, assay reagents and medicaments
XX
PS Disclosure; Fig 25; 80pp; English.
XX
CC The present DNA sequence encodes the 40 kDa subunit of cytotoxic
CC lymphocyte maturation factor (CLMF) protein. The natural CLMF protein is
CC a 75 kDa heterodimer, where the 2 subunits (see AAT77850 for the 35 kDa
CC subunit) are bonded together via one or more disulphide bonds. CLMF
CC protein is produced by a human B lymphoblastoid cell line. CLMF
CC synergistically induces, in the presence of interleukin-2, the cytolytic
CC activity of lymphokine activated killer cells. CLMF is also capable of
CC stimulating T-cell growth. A novel antibody directed against an epitope
CC of CLMF protein (see AAM24237 for epitope) is capable of neutralising
CC and/or inhibiting CLMF bioactivity. This antibody may be used as an
CC affinity ligand for purifying the CLMF protein. The antibody can also
CC be used as an assay reagent for detecting the CLMF protein, and as a
CC medicament for selectively blocking the proliferation and activation of
CC cytotoxic T cells, e.g. in transplantation.
XX
SQ Sequence 1397 BP; 390 A; 307 C; 333 G; 367 T; 0 other;

Query Match 72.8%; Score 18.2; DB 18; Length 1397;
Best Local Similarity 87.0%; Pred. No. 29;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 attctcgtcaggtgcaggtcagc 25
   || |||| ||||| ||||| |||
Db 783 attctcgcaggtgcaggtcagc 805

RESULT 19
AAT97969
ID AAT97969 standard; DNA; 1397 BP.
AC AAT97969;
XX
DT 18-MAR-1998 (first entry)
XX
DE 40 kDa subunit of human cytotoxic lymphocyte maturation factor DNA.
XX
KW Cytotoxic lymphocyte maturation factor; CLMF; 40 kDa subunit;
KW Interleukin-2; T-cell growth; antibody; lymphokine activated killer cell;
KW CLMF bioactivity inhibition; affinity ligand; transplantation;
KW cytotoxic T cell proliferation; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT CDS 41..1027
XX FT /*tag= a
XX
XX EP790308-A1.
XX
XX 20-AUG-1997.
PD

```

```

XX 09-DEC-1990; 90EP-0123670.
XX
XX 27-AUG-1990; 90US-0572284.
PR 22-DEC-1989; 89US-0455708.
PR 09-MAY-1990; 90US-0520935.
XX
PA (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI Chizzoneite RA, Gately MK, Gubler UA, Hulmes JD;
PI Pan YE, Podlanski FJ, Stern AS;
XX
DR WPI: 1997-404733/38.
DR P-PSDB; AAM36939.
XX
PT Sub-unit of cytotoxic lymphocyte maturation factor protein - useful
PT for producing protein and as antitumour agent
XX
PS Disclosure; Fig 25; 83pp; English.
XX
CC The present DNA sequence encodes a 40 kDa subunit of cytotoxic
CC lymphocyte maturation factor (CLMF) protein. The natural CLMF protein is
CC a 75 kDa heterodimer, where the 2 subunits (see AAT97970 for the 35 kDa
CC subunit) are bonded together via one or more disulphide bonds. CLMF
CC protein is produced by a human B lymphoblastoid cell line. CLMF
CC synergistically induces, in the presence of interleukin-2, the cytolytic
CC activity of lymphokine activated killer cells. CLMF is also capable of
CC stimulating T-cell growth. The subunits may be used for preparing
CC the CLMF protein and for preparing medicaments for antitumour therapy.
XX
SQ Sequence 1397 BP; 390 A; 307 C; 333 G; 367 T; 0 other;

Query Match 72.8%; Score 18.2; DB 18; Length 1397;
Best Local Similarity 87.0%; Pred. No. 29;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 attctcgtcaggtgcaggtcagc 25
   || |||| ||||| ||||| |||
Db 783 attctcgcaggtgcaggtcagc 805

RESULT 20
AAV37767
ID AAV37767 standard; CDNA; 1397 BP.
AC AAV37767;
XX
DT 17-SEP-1998 (first entry)
XX
DE Cytotoxic lymphocyte maturation factor 40 kD subunit encoding cDNA.
XX
KW Cytotoxic lymphocyte maturation factor; CLMF; cytokine; synergize;
KW Interleukin 2; IL-2; human; lymphoblastoid cell; detection; purifying;
KW proliferation; cytotoxic T cell; transplantation; antibody; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT CDS 41..1027
XX FT /*tag= a
XX FT /product= "CLMF 40 kD subunit"
XX
XX US780597-A.
XX
XX 14-JUL-1998.
XX
XX 02-JUN-1995; 95US-0460061.
XX
XX 24-MAR-1992; 92US-0857023.
PR 22-DEC-1989; 89US-0455708.
PR 09-MAY-1990; 90US-0520935.
PR 27-AUG-1990; 90US-0572284.

```


XX 17-FEB-2000.
 PD 06-AUG-1999: 99WO-US17777.
 XX
 PF 07-AUG-1998: 98US-0160065.
 PR 01-SEP-1998: 98US-0098703.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Hillman JL, Yue H, Lal P, Tang YT, Gorgone GA, Guejler KJ;
 PI Corley NC, Baughn MR;
 XX
 DR WPI: 2000-205710/18.
 DR P-PSDB: AAY69990.
 XX
 PT New human receptor-associated proteins (HRAP) useful for the diagnosis,
 PT treatment and prevention of cell proliferative, autoimmune,
 PT inflammatory, reproductive, cardiovascular, and gastrointestinal
 PT disorders -
 XX
 PS Claim 9: Pages 91-92; 99pp; English.
 XX
 CC The present sequence is a cDNA encoding human receptor-associated protein
 CC (HRAP) from incyte clone 3407686 obtained from PROST0508 cDNA library.
 CC This sequence is expressed in reproductive, nervous and
 CC and cardiovascular tissues. HRAP has cytoskeletal, immunomodulatory,
 CC antiinflammatory, cardiant, antiarteriosclerotic, hepatotropic,
 CC antithrombotic, antirheumatic, osteopathic, antiallergic, antianaemic,
 CC antasthmatic, antidiabetic, dermatological and neuroprotective
 CC activities. The present sequence is useful in the diagnosis, treatment
 CC and prevention of disorders associated with HRAP expression, especially
 CC cell proliferative, autoimmune/inflammatory, reproductive,
 CC cardiovascular and gastrointestinal disorders (e.g. atherosclerosis,
 CC cirrhosis, leukaemia, cancer, AIDS, arthritis, allergies, anaemia,
 CC asthma, dermatitis, diabetes, osteoporosis, multiple sclerosis and
 CC irritable bowel syndrome).
 XX
 SQ Sequence 2096 BP; 422 A; 614 C; 572 G; 488 T; 0 other;

Query Match 72.8%; Score 18.2; DB 21; Length 2096;
 Best Local Similarity 87.0%; Pred. No. 30;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 gatcctgcaggtgcagtcagc 24
 || ||||| ||||| |||||
 Db 73 GAGCCTCCGCGTCCGTCAG 51

RESULT 23
 AAA92406
 ID AAA92406 standard; cDNA: 2300 BP.
 XX
 AC AAA92406;
 XX
 DT 15-JAN-2001 (first entry)
 XX
 DE Human interleukin 2 encoding p35 and p40 proteins.
 XX
 KW Human; interleukin 2; IL-2; p35; p40; immune response; tumour;
 KW infectious disease; bacteria; virus; parasite; ss.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH 19..678
 FT /*tag= a
 FT /product= "p35"
 FT 1304..2290
 FT /*tag= b
 FT /product= "p40"
 XX

PN CN1258740-A.
 XX
 PD 05-JUL-2000.
 XX
 PF 31-DEC-1998; 98CN-0126748.
 XX
 PR 31-DEC-1998; 98CN-0126748.
 XX
 PA (HUAC-) HUACHEN BIOLOGICAL TECHNOLOGY INST SHANG.
 XX
 PI Cao X, Zhang W, Ju D;
 PI
 XX
 DR WPI: 2000-556956/52.
 DR P-PSDB: AAB23795, AAB23796.
 XX
 PT Recombinant adenovirus expressing human interleukin-12 and its
 PT preparation and use -
 PT
 XX
 PS Example 3; Fig 4; 27pp; Chinese.
 XX
 CC The present invention describes a recombinant adenovirus and its
 CC preparation and use, as well as medicine composite containing the
 CC recombinant adenovirus. The recombinant adenovirus is used clinically,
 CC has the advantages of low cost, easy use and high efficiency and is
 CC especially suitable for immune treatment of tumour and infectious
 CC diseases caused by bacteria, virus and parasite. The present sequence
 CC represents human interleukin 2 (IL-2) cDNA which encodes the p35 and
 CC p40 proteins, which is used in an example from the present invention.
 XX
 SQ Sequence 2300 BP; 598 A; 591 C; 571 G; 540 T; 0 other;

Query Match 72.8%; Score 18.2; DB 21; Length 2300;
 Best Local Similarity 87.0%; Pred. No. 30;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 atcctgcaggtgcagtcagc 25
 || ||||| ||||| |||||
 Db 2046 atctcgcaggtgcagtcagc 2068

RESULT 24
 AAQ23586
 ID AAQ23586 standard; cDNA: 2362 BP.
 XX
 AC AAQ23586;
 XX
 DT 16-SEP-1992 (first entry)
 XX
 DE Natural killer cell stimulatory factor 40kd subunit.
 XX
 KW NKSF; cytokine; gamma interferon; IFN; peripheral blood lymphocytes;
 KW IL-2; GM-CSF; granulocyte macrophage colony stimulating factor; PBL;
 ds.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH 33..1019
 FT /*tag= a
 FT /product= pro 40 kd subunit of NKSF
 FT
 XX
 PN WO9205256-A.
 XX
 PD 02-APR-1992.
 PD
 XX
 PF 04-SEP-1991; 91WO-US06332.
 XX
 PR 18-SEP-1990; 90US-0584941.
 XX
 PA (GENE-) GENETICS INST INC.
 PA (WIST-) WISTAR INST.
 XX

PI Clark S C, Hewick R, Kobayashi M, Perussia B, Trinchieri G;
 PI Wong G G;
 XX
 DR WPI: 1992-132124/16.
 DR P-PSDB; AAR22769.
 XX
 PT New natural killer cell stimulatory factor - useful as an
 PT immunostimulant for inducing gamma-interferon and GM-CSF prodn.
 PT in human peripheral blood lymphocytes
 XX
 PS Claim 12; Page 23; 79pp; English.
 XX
 CC NKSF is capable of inducing the production of gamma interferon in
 CC human peripheral blood lymphocytes. It has subunits of 40 (AAR22769)
 CC and 30-35 kD (AAR23729). It is claimed that NKSF is useful in the
 CC treatment of bacterial and viral infections e.g. AIDS.
 CC Degenerate probes were designed from tryptic fragments of NKSF
 CC and used to screen a cDNA library made from poly A RNA from PdHu
 CC induced 8866 cells to identify a cDNA encoding the small subunit
 CC of NKSF. cDNA prepared from 8866 poly A RNA was cloned into
 CC lambdaZAP and screened with a fragment within pNK-6 (A prior
 CC partial clone containing the N-terminal non coding region of the
 CC 40kD subunit). The library was then screened with probes designed
 CC from tryptic peptides as above, to yield pNK162. This was ligated
 CC together with pNK-6 to form pNK40-4 which is believed to contain
 CC the full length cDNA for the 40 kD NKSF subunit.
 CC See also AAR22769, AAO24371, AAR23729.
 CC
 XX
 SQ Sequence 2362 BP; 717 A; 482 C; 543 G; 620 T; 0 other;

Query Match 72.8%; Score 18.2; DB 13; Length 2362;
 Best Local Similarity 87.0%; Pred. No. 30;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 atctctgcagtgagtcagc 25
 || |||| ||||| |||||
 Db 775 atctctgcagtgagtcagc 797

RESULT 25
 AAT48098
 ID AAT48098 standard; cDNA: 2362 BP.
 XX
 AC AAT48098;
 XX
 DT 21-MAR-1997 (first entry)
 XX
 DE Human interleukin-12 40 kD subunit gene.
 XX
 KW Human; interleukin 12; IL-12; heterodimeric complex; immunogenic; ds;
 KW Leishmania; antigen; leishmaniasis; cell-mediated immunity; adjuvant;
 KW TH1 helper cell; TH2 helper cell; alum; cytokine; bacterium; cancer;
 KW vaccine; cell surface; membrane; glycoprotein; antigen presenting cell.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 33..1019
 FT /*tag= a
 FT /*product= interleukin-12 40 kD subunit
 XX
 PN US5571515-A.
 XX
 PD 05-NOV-1996.
 XX
 PF 18-APR-1994; 94US-0229282.
 XX
 PR 17-JUN-1994; 94US-0265087.
 PR 18-APR-1994; 94US-0229282.
 XX
 PA (UYPF-) UNIV PENNSYLVANIA.
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.

XX
 PI Scott P, Trinchieri G;
 XX
 DR WPI: 1996-505347/50.
 DR P-PSDB; AAW07398.
 XX
 PT Immunogenic compsn. to improve cell mediated immune response -
 PT contains soluble leishmania antigen and interleukin-12 as adjuvant
 XX
 PS Disclosure; Column 17-22; 22pp; English.
 XX
 CC This is the nucleotide sequence encoding the human interleukin (IL)-12
 CC 40 kD subunit. IL-12 is a heterodimeric complex composed of the 40 kD
 CC subunit and a 30 kD subunit (AAW07399). The complex is used in a novel
 CC immunogenic composition comprising a soluble leishmania antigen with
 CC IL-12, for protection against leishmaniasis. The addition of IL-12
 CC improves cell-mediated immunity by inducing TH1 helper cells (as opposed
 CC to TH2 cells which are induced by alum adjuvant) and does not cause
 CC uncontrolled release of other cytokines (in contrast to bacterial
 CC adjuvants). IL-12 can also be used as a cancer vaccine by association
 CC with the protein B7, a soluble, cell-surface (membrane)-bound
 CC glycoprotein which is expressed in antigen presenting cells.
 CC
 XX
 SQ Sequence 2362 BP; 717 A; 483 C; 542 G; 620 T; 0 other;

Query Match 72.8%; Score 18.2; DB 17; Length 2362;
 Best Local Similarity 87.0%; Pred. No. 30;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 atctctgcagtgagtcagc 25
 || |||| ||||| |||||
 Db 775 atctctgcagtgagtcagc 797

RESULT 26
 AAV42535
 ID AAV42535 standard; cDNA: 2362 BP.
 XX
 AC AAV42535;
 XX
 DT 09-NOV-1998 (first entry)
 XX
 DE Human interleukin-13 40 kDa subunit cDNA.
 XX
 KW Interleukin-13; IL-13; human; cytokine; adjuvant; vaccine;
 KW tetanus toxoid; HIV-1; Mycobacterium tuberculosis; influenza A;
 KW hepatitis A; hepatitis B; hepatitis C; leprosy; leishmaniasis;
 KW malaria; AIDS; cancer; tumour; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 33..1019
 FT /*tag= a
 FT /*product= interleukin-13 40 kDa subunit
 XX
 PN WO9831384-A1.
 XX
 PD 23-JUL-1998.
 XX
 PF 22-JAN-1998; 98WO-US01500.
 XX
 PR 22-JAN-1997; 97US-0036342.
 XX
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
 PA Montaner LJ;
 PI WPI: 1998-413812/35.
 DR P-PSDB; AAW62992.
 XX
 PT Enhancing immune response to antigen using interleukin-13 - useful
 PT in vaccine production, especially for antigens not responsive to

PT known adjuvants e.g. Tetanus Toxoid and to treat e.g. HIV or cancer
 XX
 PS Disclosure: Page 53-55; 95pp; English.

CC This cDNA sequence includes an open reading frame coding for the
 CC 40 kDa subunit (see AAM62992) of human Interleukin-13 (IL-13). A
 CC claimed method for enhancing the immune response to a selected
 CC antigen in a host comprises administering an effective amount of
 CC IL-13 or a biologically active fragment (IL-13'), simultaneously
 CC or sequentially with the antigen. Also claimed are: (1) a method
 CC for enhancing the adjuvant effect of Interleukin-12 (IL-12) on an
 CC antigen in a host by similarly administering IL-13 with the antigen
 CC and IL-12; (2) vaccine compositions comprising a selected antigen
 CC and IL-13, optionally with a second adjuvant; and (3) use of IL-13
 CC to prepare medicaments for administration as above. The method is
 CC used to enhance the immune response of a host to a particular
 CC antigen, especially when the IL-13 is administered in medicaments as
 CC in (2), optionally further comprising IL-12 or functional fragments
 CC (especially in which the antigen is Tetanus Toxoid (TT), influenza
 CC A, HIV-1 p24, HIV-1 nef or Mycobacterium tuberculosis) (claimed).
 CC IL-13 was shown to be an effective adjuvant in both healthy or
 CC pathogen-infected mammals (e.g. HIV-infected humans) (claimed).
 CC The method and vaccines/medicaments of (1) and (2) may be used to
 CC treat viral infections (e.g. HIV, Influenza A, Hepatitis A, B or C
 CC etc.), bacterial infections (e.g. leprosy or tuberculosis),
 CC parasitic infections (e.g. leishmaniasis or malaria) and certain
 CC cancers and solid tumours. Use of IL-13 may be preferable to use
 CC of other adjuvants since it is relatively stable in vivo (unlike
 CC IL-2), has a direct action on antigen presenting cells (unlike
 CC IL-12), and unlike bacterial adjuvants does not induce
 CC pro-inflammatory cytokines. IL-13 can also adjuvant antigens not
 CC responsive to known adjuvants (e.g. TT). IL-13 nucleic acids can
 CC be used as adjuvants, operably linked to sequences encoding
 CC IL-2, IL-12 and/or the antigen.

XX
 SQ Sequence 2362 BP; 717 A; 483 C; 542 G; 620 T; 0 other:

Query Match 72.8%; Score 18.2; DB 19; Length 2362;
 Best Local Similarity 87.0%; Pred. No. 30;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 atctcgcagcgtgcagtcagc 25
 ||||| ||||| ||||| |||||
 Db 775 attctgcgcagcgtgcagtcagc 797

RESULT 27
 AAV15227
 ID AAV15227 standard; cDNA; 2362 BP.
 XX
 AC AAV15227;

XX 27-MAY-1998 (first entry)

DE Heterodimeric human interleukin 12 subunit 1 encoding cDNA.

KM Heterodimeric; human; interleukin 12; IL-12; vaccine; antigen;
 KM pathogen; immune response; microorganism; cancer; ds.

XX Homo sapiens.

XX Key Location/Qualifiers
 FT CDS 33..1019
 FT /tag= a
 FT /product= "heterodimeric human interleukin 12 subunit"

XX US5723127-A.

XX 03-MAR-1998.

XX 25-MAR-1996; 96US-0621493.
 XX

PR 17-JUN-1994; 94US-0265087.
 PR 18-APR-1994; 94US-0229282.
 PR 25-MAR-1996; 96US-0621493.

XX
 PA (UYPE-) UNIV PENNSYLVANIA.
 PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.

XX Scott P, Trinchieri G;

XX WPI; 1998-205996/18.
 DR P-PSDB; AAM44324.

PT Enhancing immune response to vaccine - comprises co-administering
 PT interleukin-12 with antigenic composition

PS Disclosure: Column 17-22; 23pp; English.

CC The present sequence encodes a heterodimeric human interleukin 12
 CC (IL-12) subunit (putatively the 40 kD subunit) shown in the present
 CC specification. The present invention describes a method for enhancing
 CC the immune response elicited by an antigenic composition. The method
 CC comprises administering IL-12 simultaneously or sequentially with the
 CC antigenic composition. The IL-12 can be used in therapeutic
 CC compositions, e.g. as an adjuvant in vaccines against pathogenic
 CC bacteria or viruses, especially in an amount of 0.1 mu g to 0.5 mg per
 CC 25 mu g of antigen, or in cancer vaccines.

XX
 SQ Sequence 2362 BP; 717 A; 483 C; 542 G; 620 T; 0 other:

Query Match 72.8%; Score 18.2; DB 19; Length 2362;
 Best Local Similarity 87.0%; Pred. No. 30;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 atctcgcagcgtgcagtcagc 25
 ||||| ||||| ||||| |||||
 Db 775 attctgcgcagcgtgcagtcagc 797

RESULT 28
 AA228862
 ID AA228862 standard; cDNA; 2362 BP.
 XX
 AC AA228862;

XX 01-FEB-2000 (first entry)

DE Human interleukin-12 40 kD subunit gene.

KM Human; interleukin-12; IL-12; antigen; cancer; immune response; ds.

XX Homo sapiens.

XX US5976539-A.

XX 02-NOV-1999.

XX 06-NOV-1997; 97US-0965688.

XX 17-JUN-1994; 94US-0265087.

XX 25-MAR-1996; 96US-0621493.

XX 18-APR-1994; 94US-0229282.

XX (UYPE-) UNIV PENNSYLVANIA.
 XX (WIST-) WISTAR INST ANATOMY & BIOLOGY.

XX Trinchieri G, Scott P;

XX WPI; 2000-021927/02.

XX P-PSDB; AAY44181.
 XX A new anticancer vaccine -
 XX

KW IL-12 p40 subunit; gene therapy; tumour; leukaemia; Flexi-12; ss.
XX Synthetic.
OS Homo sapiens.
XX US5994104-A.
XX 30-NOV-1999.
XX
XX 08-NOV-1996; 96US-0751767.
XX
XX 08-NOV-1996; 96US-0751767.
XX
XX (UNLO) ROYAL FREE HOSPITAL SCHOOL MED.
XX
XX Anderson RJ, Prentice HG, MacDonald ID;
XX WPI: 2000-038261/03.
XX P-PSDB: AAY54921.
XX
XX Nucleic acid constructs encoding interleukin-12 fusion proteins useful
XX for treating leukemia and other cancers -
XX
XX Example; Fig 11; 73pp; English.
XX
XX This sequence encodes the human interleukin-12 fusion protein sequence,
XX designated Flexi-12. The invention relates to an isolated nucleic acid
XX construct (I) comprising a region encoding an interleukin-12 (IL-12)
XX fusion protein (comprising an IL-12 p35 subunit, an IL-12 p40 subunit and
XX a linker peptide (joining the subunits)) and a region encoding a B7
XX protein. (I) may be used to produce IL-12 fusion proteins according to
XX standard recombinant DNA methodologies. The fusion proteins may be
XX produced either in vitro (in a fermentation culture or in vivo as part of
XX a gene therapy protocol (in this case (I) is used to transform a patient's
XX cells, which then secrete the functional polypeptide to supplement the
XX patient's own production of IL-12 or to rectify mutations which lead to
XX the expression of inactive polypeptides). The fusion proteins produced in
XX this way may be used to treat any disease which responds to IL-12 such as
XX tumours (both solid and dispersed of the kidney, breast, colon, ovarian
XX and cervical tumours and melanomas) and in particular, tumours of the
XX blood such as leukaemia. Alternatively, the polypeptides may be used as
XX agonists in the production of antibodies to IL-12 and to assay for
XX agonists and antagonists of its activity. The antibodies and antagonists
XX may be used to inhibit the activity of IL-12. (I) may also be used
XX diagnostically as a probe which hybridizes to sequences encoding IL-12
XX and the antibodies may be used to detect the presence of IL-12
XX polypeptides in samples. They may be used diagnostically to quantitate
XX the expression of the polypeptide by patients and hence which subjects
XX may be in need of restorative therapy.
XX
XX Sequence 6139 BP; 1552 A; 1538 C; 1558 G; 1491 T; 0 other;
XX
XX
XX Query Match 72.8%; Score 18.2; DB 21; Length 6139;
XX Best Local Similarity 87.0%; Pred. No. 33;
XX Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX OY 3 atctcgtcaggtgcaggtcagc 25
XX || |||| |||| |||| ||||
XX Db 3964 atctcgtcaggtgcaggtcagc 3986
XX
XX
XX RESULT 32
XX AAC75563/c
XX ID AAC75563 standard; cDNA; 586 BP.
XX
XX AC AAC75563;
XX
XX 08-FEB-2001 (first entry)
XX
XX Human ORFX ORF1118 polynucleotide sequence SEQ ID NO:2235.
XX
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX
XX

KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
XX Homo sapiens.
XX
XX OS
XX
XX PN W0200058473-A2.
XX
XX PD 05-OCt-2000.
XX
XX PF 31-MAR-2000; 2000WO-US08621.
XX
XX PR 31-MAR-1999; 99US-0127607.
XX PR 02-APR-1999; 99US-0127636.
XX PR 05-APR-1999; 99US-0127728.
XX PR 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX PI Shinkets RA, Leach M;
XX
XX DR WPI: 2000-602362/57.
XX P-PSDB: AAB41354.
XX
XX PT Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX
XX PS Claim 5; Page 1649; 5507pp; English.
XX
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnary;
XX antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX antithyroid; and antianaemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antinflammatory disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 586 BP; 122 A; 185 C; 156 G; 123 T; 0 other;
XX
XX
XX Query Match 68.8%; Score 17.2; DB 21; Length 586;
XX Best Local Similarity 86.4%; Pred. No. 75;
XX Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX OY 3 atctcgtcaggtgcaggtcagc 24
XX |||| |||| |||| ||||
XX Db 304 ATCTCTGTCAGGTGCGGGGTAG 283
XX

CC	Result 33
CC	AAC76615/c
ID	AAC76615 standard; cDNA: 5269 BP.
XX	
AC	AAC76615;
XX	
DT	08-FEB-2001 (first entry)
XX	
DE	Human ORFX ORF2170 polynucleotide sequence SEQ ID NO:4339.
XX	
KW	Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
KW	vulnary; antiparasitic; antiparkinsonian; nootropic; neuroprotective;
KW	anticonvulsant; osteopathic; antihistitic; immunosuppressant; cardiac;
KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW	hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW	antiviral; antibacterial; antifungal; antirheumatic; antihypertoid;
KW	antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW	cholesterol ester storage; systemic lupus erythematosus; infection;
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW	bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW	thrombosis; contraceptive; ss.
XX	
OS	Homo sapiens.
XX	
PN	W0200058473-A2.
XX	
PD	05-OCT-2000.
XX	
PE	31-MAR-2000; 2000WO-US08621.
XX	
PR	31-MAR-1999; 99US-0127607.
XX	
PR	02-APR-1999; 99US-0127636.
XX	
PR	05-APR-1999; 99US-0127728.
XX	
PR	30-MAR-2000; 2000US-0540763.
XX	
PA	(CURA-) CURAGEN CORP.
XX	
PI	Shinkets RA, Leach M;
XX	
DR	WPI; 2000-602362/57.
XX	
PT	P-PSDB; AAB42406.
XX	
PT	Novel nucleic acids and peptides derived from open reading frame X,
XX	useful for treating e.g. cancers, proliferative disorders,
PT	neurodegenerative disorders and cardiovascular disease -
XX	
PS	Claim 5; Page 3533-3536; 5507pp; English.
XX	
CC	AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC	which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC	sequences have activities such as: cytosolic; hepatotropic; vulnary;
CC	antiparasitic; antiparkinsonian; nootropic; neuroprotective;
CC	osteopathic; anticonvulsant; antihistitic; immunosuppressant;
CC	immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;
CC	antidiabetic; hypotensive; dermatological; immunosuppressive;
CC	antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC	antihypertoid; antianaemic. The sequences can be used for determining
CC	the presence of or predisposition to, or preventing or treating
CC	pathological conditions associated with an ORFX-associated disorder. The
CC	nucleic acids can be used to express ORFX proteins in gene therapy
CC	vectors. The proteins and nucleic acids may be used to treat cancers,
CC	proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC	graft vs host disease, cardiovascular disease, diabetes mellitus,
CC	hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC	erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC	bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC	allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC	nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC	coagulation; to inhibit thrombosis; and as a contraceptive.

```

RESULT 35
AAF28005
ID AAF28005 standard; cDNA; 220 BP.
XX
AC AAF28005;
XX
DT 08-MAY-2001 (first entry)
XX
DE Human TR13 coding sequence fragment SEQ ID NO: 14.
XX
KW Human; tumour necrosis factor receptor; TR13; TR14; infection;
KW cancer; autoimmune disease; allergy; inflammatory disease;
KW graft rejection; apoptosis; cardiovascular disease; aneurysm; ss.
XX
OS Homo sapiens.
XX
PN WO200105834-A1.
XX
PD 25-JAN-2001.
XX
PF 14-JUL-2000; 2000WO-US19343.
XX
PR 16-JUL-1999; 99US-0144087.
PR 18-AUG-1999; 99US-0149450.
PR 20-AUG-1999; 99US-0149712.
PR 10-SEP-1999; 99US-0153089.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Ni J, Young PE;
XX
DR WPI; 2001-112682/12.
XX
PT Nucleic acids encoding 2 human tumor necrosis factor receptor
PT polypeptides ((TR13) and (TR14)), useful for the prevention, diagnosis
PT and treatment of, e.g. cancers, acquired immune deficiency syndrome and
PT hypohidrotic ectodermal dysplasia -
XX
PS Claim 31; Page 382; 418pp; English.
XX
CC The present invention provides the protein and coding sequences of the
CC human tumour necrosis factor receptors TR13 and TR14. These sequences are
CC useful in the diagnosis and treatment of many diseases, including cancer,
CC autoimmune diseases, cardiovascular disorders, allergies,
CC neurodegenerative diseases, graft rejection, inflammation, aneurysms and
CC infections.
XX
SQ Sequence 220 BP; 43 A; 63 C; 57 G; 56 T; 1 other;

Query Match 66.4%; Score 16.6; DB 22; Length 220;
Best Local Similarity 82.6%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cgatctctgcaggtcagtcaca 23
   | | | | | | | | | | | | | | | |
Db 59 ccacgcgcgcaggtcagtcaca 81

RESULT 36
AAF28004
ID AAF28004 standard; cDNA; 220 BP.
XX
AC AAF28004;
XX
DT 08-MAY-2001 (first entry)
XX
DE Human TR13 coding sequence fragment SEQ ID NO: 55.
XX
KW Human; tumour necrosis factor receptor; TR13; TR14; infection;
KW cancer; autoimmune disease; allergy; inflammatory disease;

```

```

KW graft rejection; apoptosis; cardiovascular disease; aneurysm; ss.
XX
OS Homo sapiens.
XX
PN WO200105834-A1.
XX
PD 25-JAN-2001.
XX
PF 14-JUL-2000; 2000WO-US19343.
XX
PR 16-JUL-1999; 99US-0144087.
PR 18-AUG-1999; 99US-0149450.
PR 20-AUG-1999; 99US-0149712.
PR 10-SEP-1999; 99US-0153089.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Ni J, Young PE;
XX
DR WPI; 2001-112682/12.
XX
PT Nucleic acids encoding 2 human tumor necrosis factor receptor
PT polypeptides ((TR13) and (TR14)), useful for the prevention, diagnosis
PT and treatment of, e.g. cancers, acquired immune deficiency syndrome and
PT hypohidrotic ectodermal dysplasia -
XX
PS Claim 33; Page 407; 418pp; English.
XX
CC The present invention provides the protein and coding sequences of the
CC human tumour necrosis factor receptors TR13 and TR14. These sequences are
CC useful in the diagnosis and treatment of many diseases, including cancer,
CC autoimmune diseases, cardiovascular disorders, allergies,
CC neurodegenerative diseases, graft rejection, inflammation, aneurysms and
CC infections.
XX
SQ Sequence 220 BP; 43 A; 63 C; 57 G; 56 T; 1 other;

Query Match 66.4%; Score 16.6; DB 22; Length 220;
Best Local Similarity 82.6%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cgatctctgcaggtcagtcaca 23
   | | | | | | | | | | | | | | | |
Db 59 ccacgcgcgcaggtcagtcaca 81

RESULT 37
AAC98199
ID AAC98199 standard; cDNA; 625 BP.
XX
AC AAC98199;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human colon cancer antigen nucleotide sequence SEQ ID NO:209.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulneryary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotoxic; antibacterial; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder; ss.
XX
OS Homo sapiens.
XX
PN WO20005351-A1.
XX
PD 21-SEP-2000.
XX
PR 08-MAR-2000; 2000WO-US05863.

```


PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 XX WPI; 2000-587534/55.
 DR P-PSDB; AAB53442.
 XX
 XX Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer -
 XX
 PS Claim 1; Page 633; 2104pp; English.
 XX
 XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
 CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
 CC vulnary, nephrotoxic, antineoplastic and antibacterial activities, and
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins
 CC may also be used to prevent diseases such as neural disorders, immune
 CC system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, wounds, renal disorders, infectious
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 CC AAB54007 represent sequences used in the exemplification of the present
 CC invention.
 XX
 XX Sequence 625 BP; 139 A; 177 C; 143 G; 156 T; 10 other;
 SQ

Query Match 66.4%; Score 16.6; DB 21; Length 625;
 Best Local Similarity 82.6%; Pred. No. 1.4e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cgaacctcgcagtcagtcagtcga 23
 ||||| ||||| ||||| ||||| ||
 Db 464 ccatcgcgcagtcagtcagtcga 486

RESULT 38
 AAF14300
 ID AAF14300 standard; cDNA; 651 BP.
 XX
 AC AAF14300;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Aspergillus oryzae EST SEQ ID NO:6823.
 XX
 XX Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 OS Aspergillus oryzae.
 XX
 PN WO200056762-A2.
 XX
 PD 28-SEP-2000.
 XX
 PF 22-MAR-2000; 2000WO-US07781.
 XX
 PR 22-MAR-1999; 99US-0273623.
 XX
 XX (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX

PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX
 DR WPI; 2000-594572/56.
 XX
 PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 XX
 PS Claim 88; Page 2778; 3161pp; English.
 XX
 XX The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.
 XX
 XX Sequence 651 BP; 185 A; 171 C; 162 G; 133 T; 0 other;
 SQ

Query Match 66.4%; Score 16.6; DB 21; Length 651;
 Best Local Similarity 82.6%; Pred. No. 1.4e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cgaacctcgcagtcagtcagtcga 23
 ||||| ||||| ||||| ||||| ||
 Db 552 cgcctcgcagtcagtcagtcga 574

RESULT 39
 AAF71312
 ID AAF71312 standard; DNA; 681 BP.
 XX
 AC AAF71312;
 XX
 DT 30-APR-2001 (first entry)
 XX
 DE Corynebacterium glutamicum HA protein nucleotide sequence SEQ ID NO:349.
 XX
 XX Corynebacterium glutamicum; homeostasis; adaptation; HA protein;
 KW fine chemical production; organic acid; proteinogenic amino acid;
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
 KW carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;
 KW diagnosis; Corynebacterium diptheriae; genetic engineering;
 KW Brevibacterium; environmental condition; ds.
 XX
 XX Corynebacterium glutamicum.
 OS
 XX
 PN WO200100842-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000WO-1B00911.
 XX


```

FT      CDS      1..990
FT      /*tag= a
FT      /product= Interleukin-12_P40_subunit
XX      JP10036397-A.
XX      PD      10-FEB-1998.
XX      PF      08-NOV-1996; 96JP-0296789.
XX      PR      23-MAY-1996; 96JP-0128104.
XX      PR      08-NOV-1995; 95JP-0289729.
XX      PA      (TORA ) TORAY IND INC.
XX      DR      WPI: 1998-174914/16.
XX      DR      P-PSDB: AAM41791.
XX      PT      Canine interleukin 12 - comprises P40 and P35 subunits; useful in
XX      PT      veterinary medicine, e.g. antitumour, antiviral and vaccine adjuvant
XX      PT      activities are expected
XX      PS      Claim 8; Page 8; 12pp: Japanese.
XX      CC      The present sequence encodes a canine interleukin-12 (IL-12) P40
XX      CC      subunit. A canine IL-12 comprising a P40 and P35 subunit is capable
XX      CC      of inducing an antiviral activating factor and the expression of
XX      CC      class II MHC molecules in canine tumour cells, stimulating
XX      CC      proliferation of canine blastogenic lymphocytes and activating
XX      CC      canine leukocytes to inhibit canine tumour cells. The canine IL-12
XX      CC      can be used in veterinary medicines, e.g. antitumour, antiviral and
XX      CC      vaccine adjuvant activities are expected.
XX      SQ      Sequence 990 BP; 278 A; 243 C; 251 G; 218 T; 0 other;

Query Match      66.4%; Score 16.6; DB 19; Length 990;
Best Local Similarity 82.6%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy      3 atctcgtcagtgagtcagc 25
      || |||| || |||| |||| ||||
Db      746 atctcgtcagtgagtcagc 768

RESULT 42
AAV13817
ID      AAV13817 standard; CDNA to mRNA; 990 BP.
XX      AC      AAV13817;
XX      DT      14-MAY-1998 (first entry)
XX      DE      Canine IL-12 P40 subunit CDNA.
XX      KM      Canine; interleukin-12 P40 subunit; IL-12 P40 subunit; antitumour;
XX      KM      antiviral; vaccine adjuvant; ds.
XX      OS      Canis sp.
XX      FH      Key
XX      FT      1..990
XX      FT      CDS      Location/Qualifiers
XX      FT      /*tag= a
XX      FT      /product= Interleukin-12_P40_subunit
XX      PN      JP10036397-A.
XX      PD      10-FEB-1998.
XX      PF      08-NOV-1996; 96JP-0296789.
XX      PR      23-MAY-1996; 96JP-0128104.
XX      PR      08-NOV-1995; 95JP-0289729.

```

```

XX      PA      (TORA ) TORAY IND INC.
XX      DR      WPI: 1998-174914/16.
XX      DR      P-PSDB: AAM41791.
XX      PT      Canine interleukin 12 - comprises P40 and P35 subunits; useful in
XX      PT      veterinary medicine, e.g. antitumour, antiviral and vaccine adjuvant
XX      PT      activities are expected
XX      PS      Claim 10; Pages 10-11; 12pp: Japanese.
XX      CC      The present sequence encodes a canine interleukin-12 (IL-12) P40
XX      CC      subunit. A canine IL-12 comprising a P40 and P35 subunit is capable
XX      CC      of inducing an antiviral activating factor and the expression of
XX      CC      class II MHC molecules in canine tumour cells, stimulating
XX      CC      proliferation of canine blastogenic lymphocytes and activating
XX      CC      canine leukocytes to inhibit canine tumour cells. The canine IL-12
XX      CC      can be used in veterinary medicines, e.g. antitumour, antiviral and
XX      CC      vaccine adjuvant activities are expected.
XX      SQ      Sequence 990 BP; 279 A; 244 C; 249 G; 218 T; 0 other;

Query Match      66.4%; Score 16.6; DB 19; Length 990;
Best Local Similarity 82.6%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy      3 atctcgtcagtgagtcagc 25
      || |||| || |||| |||| ||||
Db      746 atctcgtcagtgagtcagc 768

RESULT 43
AAK35615
ID      AAK35615 standard; CDNA to mRNA; 990 BP.
XX      AC      AAK35615;
XX      DT      09-JUL-1999 (first entry)
XX      DE      Nucleic acid encoding canine Interleukin-12 (IL-12).
XX      KM      Interleukin-12; IL-12; dog; cat; immune disease; CaII12; heterodimer;
XX      KM      tumour; skin disease; infectious disease; allergic disease; ds.
XX      OS      Canis sp.
XX      FH      Key
XX      FT      1..990
XX      FT      CDS      Location/Qualifiers
XX      FT      /*tag= a
XX      PN      JP1106350-A.
XX      PD      20-APR-1999.
XX      PF      15-MAY-1998; 98JP-0133345.
XX      PR      07-AUG-1997; 97JP-0213755.
XX      PR      16-MAY-1997; 97JP-0127690.
XX      PA      (TORA ) TORAY IND INC.
XX      DR      WPI: 1999-308068/26.
XX      DR      P-PSDB: AAT02339.
XX      PT      A prevention and treating agent containing interleukin 12 (CaII12) -
XX      PT      for prevention and treatment of dog and cat immune diseases
XX      PS      Claim 1; Page 12; 16pp: Japanese.
XX      CC      The present sequence encodes canine interleukin-12 (IL-12). The
XX      CC      specification describes a method for the prevention and treatment

```

CC of dog and cat immune diseases. The treatment used an agent comprising
 CC dog IL-12 (CaIL12) proteins to form a heterodimer. The agent is
 CC useful for preventing and treating dog and cat immune diseases,
 CC including tumours, skin diseases, infectious diseases and allergic
 CC diseases.

Sequence 990 BP; 278 A; 243 C; 251 G; 218 T; 0 other;

Query Match 66.4%; Score 16.6; DB 20; Length 990;
 Best Local Similarity 82.6%; Pred. No. 1.5e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

3 atctcgtcagtgagtcagc 25
 Db 746 atctcgtcagtgagtcagc 768

RESULT 44
 ID AAX35625 standard; cDNA to mRNA; 990 BP.

AC AAX35625;

DT 09-JUN-1999 (first entry)

DE Nucleic acid encoding canine Interleukin-12 (IL-12).

KW Interleukin-12; IL-12; dog; cat; immune disease; CaIL12; heterodimer;
 KW tumour; skin disease; infectious disease; allergic disease; ds.

OS Canis sp.

FH Key Location/Qualifiers
 FT CDS 1..990
 FT /*tag= a

PN JP11106350-A.

PD 20-APR-1999.

PE 15-MAY-1998; 98JP-0133345.

PR 07-AUG-1997; 97JP-0213755.
 PR 16-MAY-1997; 97JP-0127690.

PA (TORA) TORAY IND INC.

DR WPI: 1999-308068/26.

DR P-PSDB; AAY02342.

PT A prevention and treating agent containing interleukin 12 (CaIL12) -
 PT for prevention and treatment of dog and cat immune diseases

PS Claim 1; Page 14-15; 16pp; Japanese.

CC The present sequence encodes canine interleukin-12 (IL-12). The
 CC specification describes a method for the prevention and treatment
 CC of dog and cat immune diseases. The treatment used an agent comprising
 CC dog IL-12 (CaIL12) proteins to form a heterodimer. The agent is
 CC useful for preventing and treating dog and cat immune diseases,
 CC including tumours, skin diseases, infectious diseases and allergic
 CC diseases.

Sequence 990 BP; 279 A; 244 C; 249 G; 218 T; 0 other;

Query Match 66.4%; Score 16.6; DB 20; Length 990;
 Best Local Similarity 82.6%; Pred. No. 1.5e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

3 atctcgtcagtgagtcagc 25
 Db 746 atctcgtcagtgagtcagc 768

Db 746 atctcgtcagtgagtcagc 768

RESULT 45

ID AAX03162 standard; cDNA to mRNA; 990 BP.

AC AAX03162;

DT 30-MAR-1999 (first entry)

DE Nucleic acid encoding canine Interleukin 12.

KW Canine; interleukin 12; IL-12; feline; immunological disease; tumour;
 KW skin disease; viral infection; allergic disease; breast tumour;
 KW oesinophilic granuloma; epidermoid tumour; skin tumour; lipoma;
 KW othematoma; pneumoedema; skin soft pedicled soft tumour; anal tumour;
 KW otitis externa; dermatitis; eczema; fungal skin disease; pyoderma;
 KW allergic dermatitis; nettle rash; traumatic dermatitis; hair loss;
 KW dog parvovirus infection; distemper virus; cat plaque virus infection;
 KW feline leukaemia; allergy; pollinosis; ds.

OS Canis sp.

FH Key Location/Qualifiers
 FT CDS 1..990
 FT /*tag= a
 FT /product= Interleukin_12

PN W09851327-A1.

PD 19-NOV-1998.

PE 07-MAY-1998; 98WO-JP02031.

PR 16-MAY-1997; 97JP-0127690.

PA (TORA) TORAY IND INC.

PI Okano F, Satoh M, Yamada K;

DR WPI: 1999-070100/06.

DR P-PSDB; AAW84373.

PT New therapeutic and prophylactic agents - comprise
 PT genetically-engineered canine Interleukin 12, used to treat, e.g.
 PT canine and feline immunological diseases

PS Claim 1; Page 30-32; 45pp; Japanese.

CC The present sequence encodes a canine interleukin 12 (IL-12) protein.
 CC The IL-12 protein can be used in therapeutic or prophylactic agents.
 CC The agents can be used to prevent and treat canine and feline
 CC immunological diseases including dog and cat tumours, skin diseases,
 CC viral infections and allergic diseases, especially tumours, breast
 CC tumour, oesinophilic granuloma, epidermoid tumour, skin tumour, lipoma,
 CC othematoma, pneumoedema, skin soft pedicled soft tumour and anal
 CC tumour, skin diseases, otitis externa, dermatitis, eczema, fungal
 CC diseases of the skin, pyoderma, allergic dermatitis, nettle rash,
 CC traumatic dermatitis and hair loss; infections; dog parvovirus infection
 CC and distemper virus; cat plaque virus infection and feline leukaemia, and
 CC allergic diseases, e.g. pollinosis.

Sequence 990 BP; 279 A; 244 C; 249 G; 218 T; 0 other;

Query Match 66.4%; Score 16.6; DB 20; Length 990;
 Best Local Similarity 82.6%; Pred. No. 1.5e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

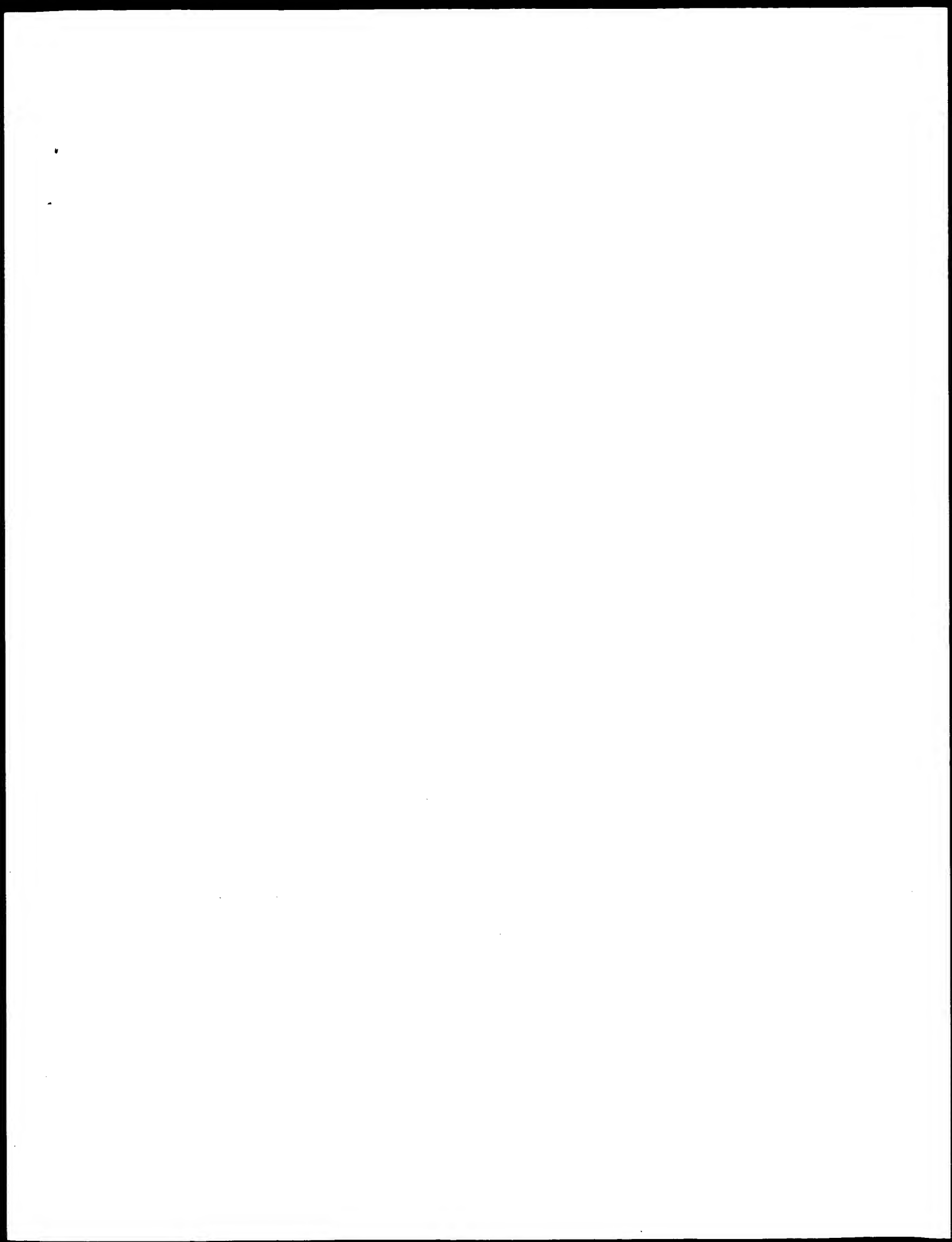
3 atctcgtcagtgagtcagc 25
 Db 746 atctcgtcagtgagtcagc 768

Wed Oct 10 07:44:57 2001

us-09-396-196f-9.std.rng

Page 23

Search completed: October 9, 2001, 11:40:03
Job time: 1918 sec



Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0077;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatcctcgtcaggtcaggtcagc 25
|||||
DB 105 CGATCCTCGTCAGGTGCAAGTCAAGC 129

RESULT 2
US-08-846-338-7
Sequence 7, Application US/08846338
Patent No. 5869719
GENERAL INFORMATION:
APPLICANT: Patton, David
TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5869719artis Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/846,338
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8587
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1038
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /product= "biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0077;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgatcctcgtcaggtcaggtcagc 25
|||||
DB 105 CGATCCTCGTCAGGTGCAAGTCAAGC 129

RESULT 3
US-08-411-768B-1
Sequence 1, Application US/08411768B
Patent No. 6083712
GENERAL INFORMATION:
APPLICANT: Olwen Birch
APPLICANT: Johann Brass
APPLICANT: Martin Fuhrmann

APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect
SOFTWARE: Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A-15/9
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 117
OTHER INFORMATION: /product= "Biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biob"
OTHER INFORMATION: /number= 1
FEATURE:
NAME/KEY: CDS
LOCATION: 2295..3050
OTHER INFORMATION: /codon_start= 2295
OTHER INFORMATION: /function= "involved in pimeloyl-coA synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "bioc"
OTHER INFORMATION: /number= 3
FEATURE:
NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3750
OTHER INFORMATION: /EC_number= 2.6.1.62
OTHER INFORMATION: /product= "DAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioA"
OTHER INFORMATION: /number= 5
OTHER INFORMATION: /standard_name= "S-adenosyl-L-methionine:8-amino-7-oxononanoate
amino transferase"
OTHER INFORMATION: /number= 5
FEATURE:
NAME/KEY: CDS


```

LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 5098
OTHER INFORMATION: /function= "unknown, involved in biotin synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ORF1"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "promoter plac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name= "promoter plac"
FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name= "bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name= "bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name= "ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name= "rho-independent"
OTHER INFORMATION: /transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "promoter plac"
OTHER INFORMATION: /evidence= EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0094;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgatcctgcaggtgcagtcagc 25
DB 221 CGATCCTGCTGAGTGCAGTCAGC 245

RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect
SOFTWARE: Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pRO30A15-9
FEATURE:
NAME/KEY: CDS
LOCATION: 1154..2308
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 1154
OTHER INFORMATION: /EC_number= 2.3.1.47
OTHER INFORMATION: /product= "KAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioF"
OTHER INFORMATION: /number= 2
OTHER INFORMATION: /standard_name= "8-Amino-7-oxononanoate synthase"
FEATURE:
NAME/KEY: CDS
LOCATION: 3043..3753
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3043
OTHER INFORMATION: /EC_number= 6.3.3.3
OTHER INFORMATION: /product= "DPR synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioD"
OTHER INFORMATION: /number= 4
OTHER INFORMATION: /standard_name= "Dethiobiotin synthase"
FEATURE:
NAME/KEY: RBS
LOCATION: 1141..1156
OTHER INFORMATION: /standard_name= "bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3030..3045
OTHER INFORMATION: /standard_name= "biob RBS"
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
```

FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-7688-6

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0094;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgaatcctgcaggtcaggtcagc 25
DB 221 CGATCCTGCAGGTGAGGTGAGGTGAGC 245

RESULT 5
US-08-186-529-1
Sequence 1, Application US/08186529
Patent No. 5573764
GENERAL INFORMATION:
APPLICANT: Sykes, Megan
APPLICANT: Wolf, Stanley F.
TITLE OF INVENTION: USE OF INTERLEUKIN-12 TO PREVENT
TITLE OF INVENTION: GRAFT-VERSUS-HOST DISEASE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Genetics Institute, Inc., Legal Affairs
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/186,529
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: McDaniel, Patricia A.
REGISTRATION NUMBER: 33,194
REFERENCE/DOCKET NUMBER: GI 5225
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-498-8401
TELEFAX: 617-876-5851
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 987 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL TYPE: Lymphoblast
CELL LINE: RPMi 8866
FEATURE:
NAME/KEY: CDS
LOCATION: 1..987
US-08-186-529-1

Query Match 72.8%; Score 18.2; DB 1; Length 987;
Best Local Similarity 87.0%; Pred. No. 8.7;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 atctcgtcaggtcaggtcagc 25
DB 743 ATTCTCGCAGGTGAGGTGAGGTGAGC 765

RESULT 6
US-08-640-386A-1
Sequence 1, Application US/08640386A
Patent No. 5756085

GENERAL INFORMATION:
APPLICANT: Sykes, Megan
APPLICANT: Wolf, Stanley F.
TITLE OF INVENTION: USE OF INTERLEUKIN-12 TO PREVENT
TITLE OF INVENTION: GRAFT-VERSUS-HOST DISEASE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Genetics Institute, Inc., Legal Affairs
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/640,386A
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: GI 5225A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-498-8224
TELEFAX: 617-876-5851

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 987 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL TYPE: Lymphoblast
CELL LINE: RPMi 8866
FEATURE:
NAME/KEY: CDS
LOCATION: 1..987
US-08-640-386A-1

Query Match 72.8%; Score 18.2; DB 1; Length 987;
Best Local Similarity 87.0%; Pred. No. 8.7;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 atctcgtcaggtcaggtcagc 25
DB 743 ATTCTCGCAGGTGAGGTGAGGTGAGC 765

RESULT 7
US-08-848-760B-24
Sequence 24, Application US/08848760B
Patent No. 6248721

GENERAL INFORMATION:
APPLICANT: Chang, Lung-Ji
TITLE OF INVENTION: Animal Model For Evaluation Of Vaccines
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: United States of America

Db 774 ATTCTCGCAGGTGAGGTGACG 796

RESULT 10

US-08-751-767A-3

Sequence 3, Application US/08751767A

Patent No. 5994104

GENERAL INFORMATION:

APPLICANT: ANDERSON, ROBERT J.

APPLICANT: GRANT, HUGH

APPLICANT: MACDONALD, IAN D.

TITLE OF INVENTION: INTERLUKIN-12 FUSION PROTEIN

NUMBER OF SEQUENCES: 80

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/751,767A

FILING DATE: 08-NOV-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: SADOFF, B.J.

REGISTRATION NUMBER: 36,663

REFERENCE/DOCKET NUMBER: 117-221

TELECOMMUNICATION INFORMATION:

TELEPHONE: 7038164091

TELEFAX: 7038164100

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1399 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 43..1026

US-08-751-767A-3

Db 785 ATTCTCGCAGGTGAGGTGACG 807

Query Match

Best Local Similarity 72.8%; Score 18.2; DB 2; Length 1399;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 785 ATTCTCGCAGGTGAGGTGACG 807

US-08-751-767A-11

Sequence 11, Application US/08751767A

Patent No. 5994104

GENERAL INFORMATION:

APPLICANT: ANDERSON, ROBERT J.

APPLICANT: GRANT, HUGH

APPLICANT: MACDONALD, IAN D.

TITLE OF INVENTION: INTERLUKIN-12 FUSION PROTEIN

NUMBER OF SEQUENCES: 80

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/751,767A

FILING DATE: 08-NOV-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: SADOFF, B.J.

REGISTRATION NUMBER: 36,663

REFERENCE/DOCKET NUMBER: 117-221

TELECOMMUNICATION INFORMATION:

TELEPHONE: 7038164091

TELEFAX: 7038164100

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1560 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1557

US-08-751-767A-11

STATE: VA

COUNTRY: USA

ZIP: 22201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/751,767A

FILING DATE: 08-NOV-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: SADOFF, B.J.

REGISTRATION NUMBER: 36,663

REFERENCE/DOCKET NUMBER: 117-221

TELECOMMUNICATION INFORMATION:

TELEPHONE: 7038164091

TELEFAX: 7038164100

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 1560 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1557

US-08-751-767A-11

Query Match

Best Local Similarity 72.8%; Score 18.2; DB 2; Length 1560;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 680 ATTCTCGCAGGTGAGGTGACG 702

US-08-751-767A-9

Sequence 9, Application US/08751767A

Patent No. 5994104

GENERAL INFORMATION:

APPLICANT: ANDERSON, ROBERT J.

APPLICANT: GRANT, HUGH

APPLICANT: MACDONALD, IAN D.

TITLE OF INVENTION: INTERLUKIN-12 FUSION PROTEIN

NUMBER OF SEQUENCES: 80

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/751,767A

FILING DATE: 08-NOV-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: SADOFF, B.J.

REGISTRATION NUMBER: 36,663

REFERENCE/DOCKET NUMBER: 117-221

TELECOMMUNICATION INFORMATION:

TELEPHONE: 7038164091

TELEFAX: 7038164100

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 1399 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 43..1026

US-08-751-767A-3

Db 785 ATTCTCGCAGGTGAGGTGACG 807

US-08-751-767A-11

Sequence 11, Application US/08751767A

Patent No. 5994104

GENERAL INFORMATION:

APPLICANT: ANDERSON, ROBERT J.

APPLICANT: GRANT, HUGH

APPLICANT: MACDONALD, IAN D.

TITLE OF INVENTION: INTERLUKIN-12 FUSION PROTEIN

NUMBER OF SEQUENCES: 80

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/751,767A

FILING DATE: 08-NOV-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: SADOFF, B.J.

REGISTRATION NUMBER: 36,663

REFERENCE/DOCKET NUMBER: 117-221

TELECOMMUNICATION INFORMATION:

TELEPHONE: 7038164091

TELEFAX: 7038164100

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 1399 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 43..1026

US-08-751-767A-3

TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1623 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1620
US-08-751-767A-9

Query Match 72.8%; Score 18.2; DB 2; Length 1623;
Best Local Similarity 87.0%; Pred. No. 9.3;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 atctctgtagtgtagtgtagc 25
|| ||||| ||||| ||||| |||||
DB 743 ATTCTCGGCGAGTGAGGTGAGC 765

RESULT 13
US-08-265-087-1
Sequence 1, Application US/08265087
Patent No. 5971515

GENERAL INFORMATION:
APPLICANT: Scott, Phillip
TITLE OF INVENTION: Compositions and Methods for Use of
TITLE OF INVENTION: IL-12 as an Adjuvant
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, PO Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/265,087
FILING DATE:

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/229,282
FILING DATE: 18-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: WST51AUSA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9206
TELEFAX: 215-540-5818

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2362 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 33..1016
US-08-265-087-1

Query Match 72.8%; Score 18.2; DB 1; Length 2362;

Best Local Similarity 87.0%; Pred. No. 9.7;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 atctctgtagtgtagtgtagc 25
|| ||||| ||||| ||||| |||||
DB 775 ATTCTCGGCGAGTGAGGTGAGC 797

RESULT 14
US-08-621-493-1
Sequence 1, Application US/08621493
Patent No. 5723127

GENERAL INFORMATION:
APPLICANT: Scott, Phillip
TITLE OF INVENTION: Compositions and Methods for Use of
TITLE OF INVENTION: IL-12 as an Adjuvant
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, PO Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/621,493
FILING DATE: 25-MAR-1996

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/265,087
FILING DATE: 17-JUN-1994
APPLICATION NUMBER: US 08/229,282
FILING DATE: 18-APR-1994

ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: WST51AUSA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9206
TELEFAX: 215-540-5818

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2362 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:

NAME/KEY: CDS
LOCATION: 33..1016
US-08-621-493-1

Query Match 72.8%; Score 18.2; DB 1; Length 2362;
Best Local Similarity 87.0%; Pred. No. 9.7;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 atctctgtagtgtagtgtagc 25
|| ||||| ||||| ||||| |||||
DB 775 ATTCTCGGCGAGTGAGGTGAGC 797

RESULT 15
US-08-965-688-1
Sequence 1, Application US/08965688
Patent No. 5976539
GENERAL INFORMATION:

```

1      ZIP: 19477
2      COMPUTER READABLE FORM:
3      MEDIUM TYPE: Floppy disk
4      COMPUTER: IBM PC compatible
5      OPERATING SYSTEM: PC-DOS/MS-DOS
6      SOFTWARE: PatentIn Release #1.0, Version #1.25
7      CURRENT APPLICATION DATA:
8      APPLICATION NUMBER: US/09/260,173
9      FILING DATE: 01-Mar-1999
10     CLASSIFICATION: <Unknown>
11     17-JUN-1994
12     18-APR-1994
13
14     PRIOR APPLICATION DATA:
15     APPLICATION NUMBER: 08/965,688
16     FILING DATE: <Unknown>
17     APPLICATION NUMBER: 08/265,087
18     FILING DATE: 17-JUN-1994
19     APPLICATION NUMBER: US 08/229,282
20     FILING DATE: 18-APR-1994
21     ATTORNEY/AGENT INFORMATION:
22     NAME: Bak, Mary E.
23     REGISTRATION NUMBER: 31,215
24     REFERENCE/DOCKET NUMBER: W515AUSA
25     TELECOMMUNICATION INFORMATION:
26     TELEPHONE: 215-540-9206
27     TELEFAX: 215-540-5818
28
29     INFORMATION FOR SEQ ID NO: 1:
30     SEQUENCE CHARACTERISTICS:
31     LENGTH: 2362 base pairs
32     TYPE: nucleic acid
33     STRANDEDNESS: double
34     TOPOLOGY: unknown
35     MOLECULE TYPE: cDNA
36     FEATURE:
37     NAME/KEY: CDS
38     LOCATION: 33..1016
39     SEQUENCE DESCRIPTION: SEQ ID NO: 1:
40     US-09-260-173-1
41
42
43     Query Match      72.8%; Score 18.2; DB 4; Length 2362;
44     Best Local Similarity 87.0%; Pred. No. 9.7;
45     Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0.
46
47     QY      3 atccctcaggtgcagtcacg 25
48             ||||| ||||| ||||| ||
49     Db      775 attctggcaggtggaggtcacg 797
50
51     RESULT 17
52     US-08-751-767A-7
53     ; Sequence 7, Application US/08751767A
54     ; Patent No. 5994104
55     ; GENERAL INFORMATION:
56     ; APPLICANT: ANDERSON, ROBERT J.
57     ; APPLICANT: GRANT, HUGH
58     ; APPLICANT: MACDONALD, IAN D.
59     ; TITLE OF INVENTION: INTERLUKIN-12 FUSION PROTEIN
60     ; NUMBER OF SEQUENCES: 80
61     ; CORRESPONDENCE ADDRESS:
62     ; ADDRESSEE: NIXON & VANDERHVE P.C.
63     ; STREET: 1100 NORTH GLEBE ROAD
64     ; CITY: ARLINGTON
65     ; STATE: VA
66     ; COUNTRY: USA
67     ; ZIP: 22201
68     ; COMPUTER READABLE FORM:
69     ; MEDIUM TYPE: Floppy disk
70     ; COMPUTER: IBM PC compatible
71     ; OPERATING SYSTEM: PC-DOS/MS-DOS
72     ; SOFTWARE: PatentIn Release #1.0, Version #1.30
73     ; CURRENT APPLICATION DATA:
74     ; APPLICATION NUMBER: US/08/751,767A

```

```

; FILING DATE: 08-NOV-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 117-221
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 7038164091
; TELEFAX: 7038164100
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6139 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3222..4841
; US-08-751-767A-7

```

```

Query Match          72.8%; Score 18.2; DB 2; Length 6139;
Best Local Similarity 87.0%; Pred. No. 11;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 3 atctcgtcaggtcagtcagc 25
    ||||| ||||| ||||| |||||
Db 3964 ATTCTCGGCGAGTCGAGTCAGC 3986

```

```

RESULT 18
US-09-079-984A-1
; Sequence 1, Application US/09079984A
; Patent No. 6231850
; GENERAL INFORMATION:
; APPLICANT: Okano, Fumiyoshi, Satoh, Masahiro,
; TITLE OF INVENTION: Canine interleukin 12, a production method
; TITLE OF INVENTION: thereof, an immune disease treatment method and preventive
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Miller & Christenbury Intellectual Property
; ADDRESSEE: Department of Schnader, Harrison, Segal and Lewis, LLP
; STREET: 1600 Market Street, 39th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/079,984A
; FILING DATE: 15-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Austin R. Miller
; REGISTRATION NUMBER: 16,602
; REFERENCE/DOCKET NUMBER: 1051-98
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-1810
; TELEFAX: (215) 568-6946
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 990 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:

```

```

; ORGANISM: Canis familiaris
; FEATURE:
; NAME/KEY: Canine IL12
; LOCATION: 1 to 987
; IDENTIFICATION METHOD: Similarity
; US-09-079-984A-1

```

```

Query Match          66.4%; Score 16.6; DB 4; Length 990;
Best Local Similarity 82.6%; Pred. No. 46;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 3 atctcgtcaggtcagtcagc 25
    ||||| ||||| ||||| |||||
Db 746 ATTCTCGGCGAGTCGAGTCAGC 768

```

```

RESULT 19
US-09-079-984A-11
; Sequence 11, Application US/09079984A
; Patent No. 6231850
; GENERAL INFORMATION:
; APPLICANT: Okano, Fumiyoshi, Satoh, Masahiro,
; TITLE OF INVENTION: Canine interleukin 12, a production method
; TITLE OF INVENTION: thereof, an immune disease treatment method and preventive
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Miller & Christenbury Intellectual Property
; ADDRESSEE: Department of Schnader, Harrison, Segal and Lewis, LLP
; STREET: 1600 Market Street, 39th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/079,984A
; FILING DATE: 15-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Austin R. Miller
; REGISTRATION NUMBER: 16,602
; REFERENCE/DOCKET NUMBER: 1051-98
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-1810
; TELEFAX: (215) 568-6946
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 990 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Canis familiaris
; FEATURE:
; NAME/KEY: Canine IL12
; LOCATION: 1 to 987
; IDENTIFICATION METHOD: Similarity
; US-09-079-984A-11

```

```

Query Match          66.4%; Score 16.6; DB 4; Length 990;
Best Local Similarity 82.6%; Pred. No. 46;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 3 atctcgtcaggtcagtcagc 25
    ||||| ||||| ||||| |||||

```


COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/429,054A
FILING DATE: 26-APR-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 842,337
FILING DATE: 20-March-1992
APPLICATION NUMBER: PCT/FR 91/00593
FILING DATE: 18-July-1991
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: French 90402094.9
FILING DATE: 20-July-1990
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Charles A. Muserlian
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 146.1137
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 661-8000
TELEFAX: (212) 661-8002
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 3509
TYPE: Nucleic acid
STRANDEDNESS: Unknown
TOPOLOGY: Unknown
MOLECULE TYPE: CDNA
US-08-429-054A-12

Query Match 64.8%; Score 16.2; DB 2; Length 3509;
Best Local Similarity 85.7%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gatccgtcagtcagtcagtc 22
||||| ||||| ||||| |||||
DB 330 GATCCAGCTCGGTCGAGAGCTC 310

RESULT 23
US-08-718-777-6/c
Sequence 6, Application US/08718777
Patent No. 5981852
GENERAL INFORMATION:
APPLICANT: Van Assche, C.
APPLICANT: Lando, D.
APPLICANT: Bruneau, J. M.
APPLICANT: Voelker, T.
APPLICANT: Gervais, M.
TITLE OF INVENTION: MODIFICATION OF SUCROSE
TITLE OF INVENTION: PHOSPHATE
TITLE OF INVENTION: SYNTHASE IN PLANTS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Barbara Rae-Venter
STREET: 260 Sheridan Avenue, Suite 440
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/718,777

FILING DATE: NOT YET ASSIGNED
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,471
FILING DATE: 27-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Barbara Rae-Venter
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: CGNE.072.0205
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)328-4400
TELEFAX: (415)328-4477
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3509 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 112..3315
US-08-718-777-6

Query Match 64.8%; Score 16.2; DB 2; Length 3509;
Best Local Similarity 85.7%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gatccgtcagtcagtcagtc 22
||||| ||||| ||||| |||||
DB 330 GATCCAGCTCGGTCGAGAGCTC 310

RESULT 24
US-09-078-862-2/c
Sequence 2, Application US/09078862
Patent No. 6091003
GENERAL INFORMATION:
APPLICANT: Nan, Guo-Ling
APPLICANT: Nagai, Chitumi
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR GENETIC
TITLE OF INVENTION: TRANSFORMATION OF PINEAPPLE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medien & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/078,862
FILING DATE: 14-MAY-1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: 09-03321
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3509 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: CDNA
US-09-078-862-2

Query Match 64.8%; Score 16.2; DB 3; Length 3509;
Best Local Similarity 85.7%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 gatctcgtcaggtgcaggtc 22
||||| ||||| ||||| |||||
DB 330 GATCCAGTCGTCGTCGTCGTCGTC 310

RESULT 25
US-09-051-341-6/c
Sequence 6, Application US/09051341
Patent No. 6124528
GENERAL INFORMATION:
APPLICANT: Shemmaker, C. K.
TITLE OF INVENTION: MODIFICATION OF SOLUBLE SOLIDS USING
SUCROSE PHOSPHATE SYNTHASE ENCODING SEQUENCE
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rae-Venter Law Group, P.C.
STREET: 260 Sheridan Avenue, Suite 440
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/051,341
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/17351
FILING DATE: 25-OCT-1996
APPLICATION NUMBER: US 08/549,016
FILING DATE: 27-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/372,200
FILING DATE: 12-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Barbara Rae-Venter, Ph.D.,
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: CGNE.110.02US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)328-4400
TELEFAX: (415)328-4477
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3509 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 112..3315
US-09-051-341-6

Query Match 64.8%; Score 16.2; DB 3; Length 3509;
Best Local Similarity 85.7%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 gatctcgtcaggtgcaggtc 22
||||| ||||| ||||| |||||
DB 330 GATCCAGTCGTCGTCGTCGTCGTC 310

RESULT 26
US-08-336-408B-3/c
Sequence 3, Application US/08336408B
Patent No. 5723329
GENERAL INFORMATION:
APPLICANT: EVANS, RONALD M.
APPLICANT: MANGELSDORF, DAVID J.
TITLE OF INVENTION: RETINOID RECEPTOR COMPOSITIONS AND
METHODS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
STREET: 444 SOUTH FLOWER STREET, SUITE 2000
CITY: LOS ANGELES
STATE: CA
COUNTRY: USA
ZIP: 90071

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/336,408B
FILING DATE: 08-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/933,453
FILING DATE: 21-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US91/00399
FILING DATE: 22-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/478,071
FILING DATE: 09-FEB-1990
ATTORNEY/AGENT INFORMATION:
NAME: REITER, STEPHEN E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9851
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-546-1995
TELEFAX: 619-546-9392
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2095 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 178..1581
US-08-336-408B-3

Query Match 63.2%; Score 15.8; DB 1; Length 2095;
Best Local Similarity 89.5%; Pred. No. 1,1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 6 ctgcgtcaggtcaggtcag 24
||||| ||||| ||||| |||||
DB 709 CTCGGCAGGTCTAGGTGAC 691

RESULT 27
PCT-US91-00399-3/c
Sequence 3, Application PC/TUS9100399
GENERAL INFORMATION:
APPLICANT: Mangelsdorf, Dr., David J.
APPLICANT: Evans Dr., Ronald M.
TITLE OF INVENTION: RETINOID RECEPTOR COMPOSITIONS AND

TITLE OF INVENTION: METHODS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: FITCH, EVEN, TABIN & FLANNERY
STREET: 135 South LaSalle Street, Suite 900
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/00399
FILING DATE: 19910122
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/478,071
FILING DATE: 09-FEB-1990
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Mr., Stephen E.
REGISTRATION NUMBER: 31192
REFERENCE/DOCKET NUMBER: 50852
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 552-1311
TELEFAX: (619) 552-0095
TELEX: 20 6566 PATLAW CGO
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2095 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: RMR MOUSE ALPHA
FEATURE:
NAME/KEY: CDS
LOCATION: join(178..1581)
PCT-US91-00399-3

Query Match 63.2%; Score 15.8; DB 5; Length 2095;
Best Local Similarity 89.5%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 ctcgtcaggtgcaggtcag 24
||||| ||||| ||||| |||||
Db 709 CTCGCGAGGTGTAGGTGAG 691

RESULT 28
5457037-1/c
; Patent No. 5457037
; APPLICANT: TOGNONI, ANGELO; CARRERA, PAOLO; CAMERINI, BARBARA;
; GALLI, GIULIANO; LUCCHESI, GIUSEPPE; GRANDI, GUIDO; DI GENNARO, CARLO
; TITLE OF INVENTION: CLONING OF THE GENE CODING THE ISOMYLASE
; ENZYME AND ITS USE IN THE PRODUCTION OF SAID ENZYME
; NUMBER OF SEQUENCES: 7
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/1,797
; FILING DATE: 08-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 749,621
; FILING DATE: 19-AUG-1991
; APPLICATION NUMBER: 224,114
; FILING DATE: 25-JUL-1988
; SEQ ID NO: 1:
; LENGTH: 3336
5457037-1

Query Match 62.4%; Score 15.6; DB 6; Length 3336;
Best Local Similarity 81.8%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 cgcctcgtcaggtgcaggtc 22
||||| ||||| ||||| |||||
Db 81 CGATCTGACAGGTGATGCAAC 60

RESULT 29
US-08-055-945-1/c
; Sequence 1, Application US/08055945
; Patent No. 5462855
; GENERAL INFORMATION:
; APPLICANT: Springer, Wolfgang; Rast, Hans George;
; TITLE OF INVENTION: A METHOD FOR THE DETECTION AND
; TITLE OF INVENTION: QUANTITATIVE DETERMINATION OF
; TITLE OF INVENTION: NITROSOMONAS STRAINS IN
; TITLE OF INVENTION: WASTEWATERS OR SOILS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SPRUNG HORN KRAMER & WOODS
; STREET: 1140 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: diskette, 5.25 inch, 1.2 MB
MEDIUM TYPE: storage
COMPUTER: NEC Powermate 1 Plus
OPERATING SYSTEM: DOS
SOFTWARE: Wordperfect 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/055,945
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/692,745
FILING DATE: No. 5462855e
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: Bayer 8194-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 391-0520
TELEFAX: (212) 382-0949
TELEX: 423092 NYP UI
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1722 bp
TYPE: Nucleotide
STRANDEDNESS: Double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-055-945-1

Query Match 61.6%; Score 15.4; DB 1; Length 1722;
Best Local Similarity 76.0%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 cgcctcgtcaggtgcaggtcag 25
||||| ||||| ||||| |||||
Db 970 CGATCTGACAGGTGATGCAAC 946

RESULT 30
US-08-850-049-122/c
; Sequence 122, Application US/08850049
; Patent No. 5965726

GENERAL INFORMATION:
APPLICANT:
APPLICANT:
TITLE OF INVENTION: METHOD OF ELIMINATING
INHIBITORY/INSTABILITY REGIONS OF mRNA
NUMBER OF SEQUENCES: 130
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/850,049
FILING DATE: 02-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/050,478
FILING DATE: 26-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/02908
FILING DATE: 29-MAR-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/858,747
FILING DATE: 27-MAR-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MORRY, MARY J.
REGISTRATION NUMBER: 34,398
REFERENCE/DOCKET NUMBER: 2026-4006051
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)758-4800
TELEFAX: (212)751-6849
INFORMATION FOR SEQ ID NO: 122:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
US-08-850-049-122

Query Match 60.8%; Score 15.2; DB 2; Length 50;
Best Local Similarity 85.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgaatcctcgatcagtcagc 20
|||||
Db 43 CGATCCTCGTACGATCAGC 24

RESULT 31
US-08-050-478-122/c
Sequence 122, Application US/08050478
Patent No. 5972596
GENERAL INFORMATION:
APPLICANT:
APPLICANT:
TITLE OF INVENTION: METHOD OF ELIMINATING
INHIBITORY/INSTABILITY REGIONS OF mRNA
NUMBER OF SEQUENCES: 130
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/050,478
FILING DATE: 26-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/02908
FILING DATE: 29-MAR-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/858,747
FILING DATE: 27-MAR-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MORRY, MARY J.
REGISTRATION NUMBER: 34,398
REFERENCE/DOCKET NUMBER: 2026-4006051
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)758-4800
TELEFAX: (212)751-6849
INFORMATION FOR SEQ ID NO: 122:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
US-08-050-478-122

Query Match 60.8%; Score 15.2; DB 2; Length 50;
Best Local Similarity 85.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgaatcctcgatcagtcagc 20
|||||
Db 43 CGATCCTCGTACGATCAGC 24

RESULT 32
US-08-676-279-42/c
Sequence 42, Application US/08676279
Patent No. 5869247
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: MACROPHAGE NUCLEOTIDE SEQUENCE
NUMBER OF SEQUENCES: 63
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/676,279
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/00095
FILING DATE: 19-JAN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9400929.7
FILING DATE: 31-OCT-1994
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 769 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-676-279-42

Query Match 60.8%; Score 15.2; DB 2; Length 769;
Best Local Similarity 85.0%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 6 tctgtcaggtcaggtcagc 25
Db 253 CTGGCAGGTTGAGGTGAGC 234

RESULT 33
US-09-426-557-5/c
; Sequence 5, Application US/09426557
; Patent No. 6232527
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; TITLE OF INVENTION: Maize Rad2/FEN-1 Orthologues and Uses
; FILE REFERENCE: 0961
; CURRENT APPLICATION NUMBER: US/09/426,557
; CURRENT FILING DATE: 1999-10-22
; EARLIER APPLICATION NUMBER: 60/112,332
; EARLIER FILING DATE: 1998-12-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 1381
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (37)...(1173)
US-09-426-557-5

Query Match 60.8%; Score 15.2; DB 4; Length 1381;
Best Local Similarity 85.0%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 tctgtcaggtcaggtcagc 23
Db 947 TCCTCATCAGGTGAGGTGCA 928

RESULT 34
US-09-426-557-1/c
; Sequence 1, Application US/09426557
; Patent No. 6232527
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; TITLE OF INVENTION: Maize Rad2/FEN-1 Orthologues and Uses
; FILE REFERENCE: 0961
; CURRENT APPLICATION NUMBER: US/09/426,557
; CURRENT FILING DATE: 1999-10-22
; EARLIER APPLICATION NUMBER: 60/112,332
; EARLIER FILING DATE: 1998-12-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1463
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (85)...(1221)
US-09-426-557-1

Query Match 60.8%; Score 15.2; DB 4; Length 1463;
Best Local Similarity 85.0%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 tctgtcaggtcaggtcagc 23
Db 995 TCCTCATCAGGTGAGGTGCA 976

RESULT 35
US-09-426-557-7/c
; Sequence 7, Application US/09426557
; Patent No. 6232527
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; TITLE OF INVENTION: Maize Rad2/FEN-1 Orthologues and Uses
; FILE REFERENCE: 0961
; CURRENT APPLICATION NUMBER: US/09/426,557
; CURRENT FILING DATE: 1999-10-22
; EARLIER APPLICATION NUMBER: 60/112,332
; EARLIER FILING DATE: 1998-12-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 1478
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (97)...(1233)
US-09-426-557-7

Query Match 60.8%; Score 15.2; DB 4; Length 1478;
Best Local Similarity 85.0%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 tctgtcaggtcaggtcagc 23
Db 1007 TCCTCATCAGGTGAGGTGCA 988

RESULT 36
US-09-426-557-3/c
; Sequence 3, Application US/09426557
; Patent No. 6232527
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; TITLE OF INVENTION: Maize Rad2/FEN-1 Orthologues and Uses
; FILE REFERENCE: 0961
; CURRENT APPLICATION NUMBER: US/09/426,557
; CURRENT FILING DATE: 1999-10-22
; EARLIER APPLICATION NUMBER: 60/112,332
; EARLIER FILING DATE: 1998-12-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1541
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (79)...(1215)
US-09-426-557-3

Query Match 60.8%; Score 15.2; DB 4; Length 1541;
Best Local Similarity 85.0%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 tctctcaggtgcaggtca 23
||||| ||||| ||||| ||
Db 989 tccctcatcaggtgcaggtcca 970

RESULT 37

US-08-528-199-2/C
; Sequence 2, Application US/08528199
; Patent No. 5763228
; GENERAL INFORMATION:
; APPLICANT: KUBOTA, Michio
; APPLICANT: TSUSAKI, Keiji
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: RECOMBINANT ENZYME FOR CONVERTING
; TITLE OF INVENTION: MALTOSE INTO TREHALOSE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/528,199
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,126
; FILING DATE: 14-SEP-1995
; APPLICATION NUMBER: US 08/485,126
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 156399/1994
; FILING DATE: 16-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: KUBOTA-6A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-737-3528
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1704 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-528-199-2

Query Match 60.8%; Score 15.2; DB 1; Length 1704;
Best Local Similarity 85.0%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 gatctctcaggtgcaggt 21
||||| ||||| ||||| ||
Db 1293 gatctctcaggtgcaggt 1274

RESULT 38
US-08-528-199-5/C
; Sequence 5, Application US/08528199
; Patent No. 5763228
; GENERAL INFORMATION:
; APPLICANT: KUBOTA, Michio

APPLICANT: TSUSAKI, Keiji
APPLICANT: SUGIMOTO, Toshiyuki
TITLE OF INVENTION: RECOMBINANT ENZYME FOR CONVERTING
TITLE OF INVENTION: MALTOSE INTO TREHALOSE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/528,199
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/485,126
FILING DATE: 14-SEP-1995
APPLICATION NUMBER: US 08/485,126
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 156399/1994
FILING DATE: 16-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: KUBOTA-6A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1704 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1704
US-08-528-199-5

Query Match 60.8%; Score 15.2; DB 1; Length 1704;
Best Local Similarity 85.0%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 gatctctcaggtgcaggt 21
||||| ||||| ||||| ||
Db 1293 gatctctcaggtgcaggt 1274

RESULT 39
US-08-323-443B-1/C
; Sequence 1, Application US/08323443B
; Patent No. 5654170
; GENERAL INFORMATION:
; APPLICANT: KLINGER, KATHERINE W.
; APPLICANT: LANDES, GREGORY M.
; APPLICANT: BURN, TIMOTHY C.
; APPLICANT: CONNORS, TIMOTHY D.
; APPLICANT: DACKOWSKI, WILLIAM R.
; APPLICANT: GERMINO, GREGORY
; APPLICANT: QIAN, PENG
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
; NUMBER OF SEQUENCES: 8

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Darby & Darby PC
;; STREET: 805 Third Avenue
;; CITY: New York
;; STATE: NY
;; COUNTRY: USA
;; ZIP: 10022
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/323,443B
;; FILING DATE: 12-OCT-1994
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Ludwig, S. Peter
;; REGISTRATION NUMBER: 25,351
;; REFERENCE/DOCKET NUMBER: 0372/0A462
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 527-7700
;; TELEFAX: (212) 753-6237
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 31571 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHEICAL: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Homo sapiens
;; IMMEDIATE SOURCE:
;; CLONE: PKD1 GENOMIC
;; US-08-323-443B-1

Query Match 60.8%; Score 15.2; DB 1; Length 31571;
Best Local Similarity 85.0%; Pred. No. 2,9e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 cctcgtcaggtgcaggtcag 24
||||| ||||| ||||| |||
Db 13451 CCTGGCAGGGGCGAGCAG 13432

RESULT 40
US-08-658-136-2/C
Sequence 2, Application US/08658136
Patent No. 6071717
GENERAL INFORMATION:
APPLICANT: KLINGER, KATHERINE W
APPLICANT: LANDES, GREGORY M
APPLICANT: BURN, TIMOTHY C
APPLICANT: CONNORS, TIMOTHY D
APPLICANT: DACKOWSKI, WILLIAM
APPLICANT: GERMINO, GREGORY
APPLICANT: QIAN, FENG
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: ONE MOUNTAIN ROAD
CITY: FRAMINGHAM
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/658,136
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: LASSEN, ELIZABETH
;; REGISTRATION NUMBER: 31,845
;; REFERENCE/DOCKET NUMBER: GEN4-17.8
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 508-872-8400
;; TELEFAX: 508-872-5415
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 53526 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; US-08-658-136-2

Query Match 60.8%; Score 15.2; DB 3; Length 53526;
Best Local Similarity 85.0%; Pred. No. 3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 cctcgtcaggtgcaggtcag 24
||||| ||||| ||||| |||
Db 14436 CCTGGCAGGGGCGAGCAG 14417

RESULT 41
US-08-658-136-1/C
Sequence 1, Application US/08658136
Patent No. 6071717
GENERAL INFORMATION:
APPLICANT: KLINGER, KATHERINE W
APPLICANT: LANDES, GREGORY M
APPLICANT: BURN, TIMOTHY C
APPLICANT: CONNORS, TIMOTHY D
APPLICANT: DACKOWSKI, WILLIAM
APPLICANT: GERMINO, GREGORY
APPLICANT: QIAN, FENG
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: ONE MOUNTAIN ROAD
CITY: FRAMINGHAM
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/658,136
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LASSEN, ELIZABETH
REGISTRATION NUMBER: 31,845
REFERENCE/DOCKET NUMBER: GEN4-17.8
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-872-8400
TELEFAX: 508-872-5415
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 53577 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-08-658-136-1

Query Match 60.8%; Score 15.2; DB 3; Length 53577;
Best Local Similarity 85.0%; Pred. No. 3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 cctgcaggtgcaggtcag 24
|||||
Db 14435 CCTCGCAGGGGCGCAGCAG 14416

RESULT 42
US-08-905-223-225/c
Sequence 225, Application US/08905223
Patent No. 6222029
GENERAL INFORMATION:
APPLICANT: Edwards, Jean-Baptiste D.
APPLICANT: Duclert, Aymeric
APPLICANT: Lacroix, Bruno
TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS
NUMBER OF SEQUENCES: 503
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 501 West Broadway
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-3505
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Win95
SOFTWARE: Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/905,223
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Israelien, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 225:
SEQUENCE CHARACTERISTICS:
LENGTH: 387 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
TISSUE TYPE: Brain
FEATURE:
NAME/KEY: other
LOCATION: complement(75..325)
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 100
OTHER INFORMATION: region 82..332
OTHER INFORMATION: id AA004751
FEATURE:
NAME/KEY: other
LOCATION: complement(88..255)
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 97
OTHER INFORMATION: region 153..320
OTHER INFORMATION: id N27443
OTHER INFORMATION: est

NAME/KEY: other
LOCATION: complement(18..105)
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 94
OTHER INFORMATION: region 304..391
OTHER INFORMATION: id N27443
FEATURE:
NAME/KEY: other
LOCATION: complement(258..325)
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 100
OTHER INFORMATION: region 81..148
OTHER INFORMATION: id N27443
FEATURE:
NAME/KEY: other
LOCATION: complement(22..325)
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 99
OTHER INFORMATION: region 80..383
OTHER INFORMATION: id AA015608
OTHER INFORMATION: est
FEATURE:
NAME/KEY: other
LOCATION: complement(78..253)
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 99
OTHER INFORMATION: region 165..340
OTHER INFORMATION: id H09727
OTHER INFORMATION: est
FEATURE:
NAME/KEY: other
LOCATION: complement(253..285)
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 96
OTHER INFORMATION: region 132..164
OTHER INFORMATION: id H09727
OTHER INFORMATION: est
FEATURE:
NAME/KEY: other
LOCATION: complement(49..276)
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 97
OTHER INFORMATION: region 133..360
OTHER INFORMATION: id AA027099
OTHER INFORMATION: est
FEATURE:
NAME/KEY: other
LOCATION: complement(269..325)
IDENTIFICATION METHOD: blastn
OTHER INFORMATION: identity 100
OTHER INFORMATION: region 83..139
OTHER INFORMATION: id AA027099
OTHER INFORMATION: est
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 139..369
IDENTIFICATION METHOD: Von Heljne matrix
OTHER INFORMATION: score 5.3
OTHER INFORMATION: seq TWVFTCLVFFCFC/LIS
US-08-905-223-225

Query Match 60.0%; Score 15; DB 4; Length 387;
Best Local Similarity 78.3%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 atccctgacgtgcaggtcagc 25
|||||
Db 274 ACCCTAGTGGGCGCAGTGAGC 252

RESULT 43
US-08-998-416-995/c
; Sequence 995, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Knechtle, Jurgen
; APPLICANT: Redischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSTYPII
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 995:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1611RP
US-08-998-416-995
Query Match 60.0%; Score 15; DB 4; Length 666;
Best Local Similarity 78.3%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 gatccctgcaggtgcaggtcag 24
||||| ||||||||| |||
Db 161 GATCGCGTCAGTGCATTCAG 139

APPLICANT: Redischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSTYPII
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 1093:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 707 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1667RP
US-08-998-416-1093
Query Match 60.0%; Score 15; DB 4; Length 707;
Best Local Similarity 78.3%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 gatccctgcaggtgcaggtcag 24
||||| ||||||||| |||
Db 161 GATCGCGTCAGTGCATTCAG 139
RESULT 45
US-08-317-880-3/c
; Sequence 3, Application US/08317880
; Patent No. 5593862
; GENERAL INFORMATION:
; APPLICANT: Hall, Linda M.
; TITLE OF INVENTION: Proteins for Cation Channel Expression
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 13:46:49 ; Search time 5930.9 Seconds
(without alignments)
39.846 Million cell updates/sec

Title: US-09-396-196F-9

Perfect score: 25
Sequence: 1 cgatcctgctcagtgccagcagc 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1022815 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estp11:*
73: em_estp12:*
74: em_estp13:*
75: em_estp14:*
76: em_estp15:*
77: em_estp16:*
78: em_estp17:*
79: em_estp18:*
80: em_estp19:*
81: em_estp110:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est33:*
111: gb_est34:*
112: gb_est35:*
113: gb_est36:*
114: gb_est37:*
115: gb_est38:*
116: gb_est39:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est110:*
191: gb_est111:*
192: gb_est112:*
193: gb_est113:*
194: gb_est114:*
195: gb_est115:*
196: gb_est116:*
197: gb_est117:*
198: gb_est118:*
199: gb_est119:*
200: gb_est120:*
201: gb_est121:*
202: gb_est122:*
203: gb_est123:*
204: gb_est124:*
205: gb_est125:*
206: gb_est126:*
207: gb_est127:*
208: gb_est128:*
209: gb_est129:*
210: gb_est130:*
211: gb_est131:*
212: gb_est132:*
213: gb_est133:*
214: gb_est134:*
215: gb_est135:*
216: gb_est136:*
217: gb_est137:*
218: gb_est138:*
219: gb_est139:*
220: gb_est140:*
221: gb_est141:*
222: gb_est142:*
223: gb_est143:*
224: gb_est144:*
225: gb_est145:*
226: gb_est146:*
227: gb_est147:*
228: gb_est148:*
229: gb_est149:*
230: gb_est150:*
231: gb_est151:*
232: gb_est152:*
233: gb_est153:*
234: gb_est154:*
235: gb_est155:*
236: gb_est156:*
237: gb_est157:*
238: gb_est158:*
239: gb_est159:*
240: gb_est160:*
241: gb_est161:*
242: gb_est162:*
243: gb_est163:*
244: gb_est164:*
245: gb_est165:*
246: gb_est166:*
247: gb_est167:*
248: gb_est168:*
249: gb_est169:*
250: gb_est170:*
251: gb_est171:*
252: gb_est172:*
253: gb_est173:*
254: gb_est174:*
255: gb_est175:*
256: gb_est176:*
257: gb_est177:*
258: gb_est178:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	18.8	75.2	543	239	AZ168021	SP_0103_B
C 2	18.2	72.8	304	107	AU100606	AU100606
C 3	18.2	72.8	354	121	AM826993	AM826993
C 4	18.2	72.8	371	112	AM143033	AM143033
C 5	18.2	72.8	461	234	AQ846280	LMJFV1_1
C 6	18.2	72.8	872	220	CNS02672	AL183824
C 7	18.2	72.8	883	106	AL556337	Tetradon
C 8	18.2	72.8	904	172	BC023733	AL556337
C 9	18.2	72.8	917	106	AL580709	BC023733
C 10	18.2	72.8	1095	172	BF979422	AL580709
C 11	17.8	71.2	494	236	AQ982558	BF979422
C 12	17.8	71.2	536	165	BE234470	RPCT-23-2
C 13	17.8	71.2	777	106	AL538687	BE234470
C 14	17.8	71.2	1032	221	CNS04400	AL538687
C 15	17.6	70.4	339	9	AA594461	AL274571
C 16	17.6	70.4	360	147	BF318689	AL274571
C 17	17.6	70.4	369	151	BF659032	AA594461
C 18	17.6	70.4	398	141	BE860402	BF659032
C 19	17.6	70.4	413	10	AA644735	BE860402
C 20	17.6	70.4	413	141	BE862745	AA644735
C 21	17.6	70.4	424	225	AQ172663	BE862745
C 22	17.6	70.4	429	227	AQ371411	AQ172663
C 23	17.6	70.4	463	141	BE850719	AQ371411
C 24	17.6	70.4	495	24	AL789420	BE850719
C 25	17.6	70.4	508	17	AL180705	AL789420
C 26	17.6	70.4	618	8	AA536738	AL180705
C 27	17.6	70.4	628	138	BE653574	AA536738
C 28	17.6	70.4	641	243	AZ426129	BE653574
C 29	17.6	70.4	721	165	BE262905	AZ426129
C 30	17.6	70.4	833	154	BC478947	BE262905
C 31	17.6	70.4	893	169	BF794272	BC478947
C 32	17.6	70.4	897	220	CNS02672	BF794272
C 33	17.6	70.4	929	147	BF313136	CNS02672
C 34	17.6	70.4	942	219	CNS0001P	BF313136
C 35	17.6	70.4	982	219	CNS000PE	CNS0001P
C 36	17.6	70.4	985	219	CNS00HWP	CNS000PE
C 37	17.6	70.4	1002	171	BF965985	CNS00HWP
C 38	17.6	70.4	1016	219	CNS00H01	BF965985
C 39	17.4	69.6	1126	140	BE795069	CNS00H01
C 40	17.4	69.6	1298	152	BC328084	BE795069
C 41	17.2	68.8	108	115	AA371554	BC328084
C 42	17.2	68.8	112	10	AA651689	AA371554
C 43	17.2	68.8	235	3	AA197041	AA651689
C 44	17.2	68.8	236	147	BF349456	AA197041
C 45	17.2	68.8	255	159	N83353	BF349456

ALIGNMENTS

RESULT 1
LOCUS AZ168021/c
DEFINITION SP_0103_B1.P05.T7A Strongylocentrotus purpuratus, purple sea urchin
clone Plate=103 Col=9 Row=L, DNA sequence.

ACCESSION AZ168021
VERSION AZ168021.1
SOURCE GSS
KEYWORDS Strongylocentrotus purpuratus.
ORGANISM Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoidae; Strongylocentrotus.

REFERENCE 1 (bases 1 to 543)
AUTHORS Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray

TITLE
JOURNAL
MEDLINE
COMMENT
Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 103 row: L column: 9
Seq primer: 17
Class: BAC ends
High quality sequence stop: 543.

FEATURES
source
1..543
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="Plate=103 Col=9 Row=L"
/clone.lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BACs.6; BAC Clones in E-Coli DH10B"

BASE COUNT 149 a 116 c 116 g 157 t 5 others
ORIGIN

Query Match
Best Local Similarity 75.2%; Score 18.8; DB 239; Length 543;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 atctctgcagtgatgcagtcag 24
Db 115 ATCTCTGCAGTGCTGTGTAAG 94

RESULT 2
LOCUS AU100606/c
DEFINITION AU100606 Sugano Homo sapiens CDNA library Homo sapiens CDNA clone
y79a3n51 similar to Homo sapiens mRNA for cystinosin, mRNA
sequence.

ACCESSION AU100606
VERSION AU100606.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 300)
AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata
H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo
K., Suyama,A. and Sugano,S.
In silico mapping of the 5'-ends of human mRNAs using full-length
enriched and 5'-end enriched cDNA libraries constructed by
oligo-capping method

JOURNAL
COMMENT
Unpublished (2001)
Contact: Yutaka Suzuki
Department of Medical Science, University of Tokyo
4-6-1, Shirokane, Minato-ku, Tokyo 108-8639, Japan
Email: yusuzuki@ms.u-tokyo.ac.jp
Suzuki,Y., Yoshimoto-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
S. Construction and characterization of a full length-enriched and
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES
source
1..300
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="y79a3n51"

BASE COUNT 51 a /clone_lib="Sugano Homo sapiens cDNA library"
 ORIGIN 93 c 97 g 51 t 8 others

Query Match 72.8%; Score 18.2; DB 107; Length 300;
 Best Local Similarity 87.0%; Pred. No. 4.7e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gatccctgtaaggtgcagtcag 24
 ||||| ||||| ||||| |||||
 Db 62 GAGCTGCGCAGGTGCGGTCAG 40

RESULT 3 354 bp mRNA EST 17-MAY-2000
 AM826993/c LOCUS
 DEFINITION fk55q12.y1 zebrafish adult retina cDNA Danio rerio cDNA 5' similar
 to TR:015846 Q15846 HYPOTHETICAL 54.2 KD PROTEIN PRECURSOR. ; mRNA
 sequence.

ACCESSION AM826993.1 GI:7920070
 VERSION AM826993
 KEYWORDS EST.
 SOURCE zebrafish.
 ORGANISM Danio rerio

REFERENCE 1 (bases 1 to 354)
 Actinopterygii; Neopterygii; Teleostei; Euteleostomi;
 Cypriniformes; Cyprinidae; Rasbortinae; Danio.
 Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Maria, M., Eddy,
 S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wyllie, T., Underwood,
 K., Stepoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,
 Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,
 Kohn, S., Shin, T., Jackson, I., Cardenas, M., McCann, R., Waterston, R.,
 and Wilson, R.

TITLE WashU Zebrafish EST Project 1998
 JOURNAL Unpublished (1998)
 COMMENT Contact: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: zbrafish@watson.wustl.edu
 Library constructed by: Susan E. Brockerhoff DNA Sequencing by:
 Washington University Genome Sequencing Center Clone distribution:
 Ressourcenzentrum Primatendatenbank, Berlin, Germany (web address:
 www.rzpd.de)
 Seq primer: T3 ET from Amersham
 High quality sequence stop: 292.

FEATURES
 source
 1..354
 /organism="Danio rerio"
 /strain="wild-type"
 /db_xref="taxon:7955"
 /clone_lib="Zebrafish adult retina cDNA"
 /sex="mixed"
 /dev_stage="1-2 years"
 /lab_host="E. coli XL1-Blue MRF" (XL1-Blue MRF)"
 /note="Vector: Lambda ZAP II (pBluescript SK-); Site_1:
 EcoRI; Site_2: SalI; This zebrafish library was
 constructed by Dr. Susan E. Brockerhoff (email:
 sbrocker@u.washington.edu) RZPD library number: 760"

BASE COUNT 83 a 102 c 81 g 87 t 1 others
 ORIGIN

Query Match 72.8%; Score 18.2; DB 121; Length 354;
 Best Local Similarity 83.3%; Pred. No. 4.7e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 cgatccctgtaaggtgcagtcag 24
 ||||| ||||| ||||| |||||
 Db 235 CGATCCTGACCACTGCGGTCAG 212

RESULT 4

LOCUS AM143033/c

DEFINITION AM143033 371 bp mRNA EST 30-OCT-1999
 EST293328 Normalized rat kidney. Bento Soares Rattus sp. cDNA clone
 R61BC08 5' end, mRNA sequence.

ACCESSION AM143033

VERSION AM143033.1 GI:6162932

KEYWORDS EST.

SOURCE Rattus sp.

ORGANISM Rattus sp.

REFERENCE 1 (bases 1 to 371)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

Lee, N.H., Glodok, A., Chandra, I., Mason, T.M., Quackenbush, J.,
 Kerlavage, A.R., and Adams, M.D.

Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat
 Gene Index

Unpublished (1998)

Contact: Lee, NH
 The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529

Fax: (301)-838-0208

Email: nhlee@tigr.org
 For clone availability, additional sequence and expression
 information related to this EST please check the TIGR Rat Gene
 Index (<http://www.tigr.org/tcbl/rat/rat.html>). To order a clone
 contact the ATCC (<http://www.atcc.org/atcc.html>).
 Seq primer: M13 Reverse.

FEATURES
 source
 1..371
 /organism="Rattus sp."
 /db_xref="taxon:10118"
 /clone="RGRC08"
 /clone_lib="Normalized rat kidney, Bento Soares"
 /note="Organ: Kidney; Vector: pT73Pac; Site_1: EcoRI;
 Site_2: NotI"

BASE COUNT 82 a 117 c 103 g 69 t
 ORIGIN

Query Match 72.8%; Score 18.2; DB 112; Length 371;
 Best Local Similarity 87.0%; Pred. No. 4.7e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gatccctgtaaggtgcagtcag 24
 ||||| ||||| ||||| |||||
 Db 212 GCTCCTCTCAGGTGCGGTCAG 190

RESULT 5

LOCUS AO846280/c

DEFINITION AO846280 461 bp DNA GSS 13-APR-2001
 LMAJFV1.Lm17b03.y1 Leishmania major FV1 random genomic library
 Leishmania major genomic clone LMAJFV1.Lm17b03 5', DNA sequence.

ACCESSION AO846280

VERSION AO846280.1 GI:6050928

KEYWORDS GSS.

SOURCE Leishmania major.

ORGANISM Leishmania major.

REFERENCE 1 (bases 1 to 461)
 Eukaryota; Eukaryota; Kinetoplastida; Trypanosomatidae;
 Leishmania.

Akopyants, N.S., Clifton, S.W., Martin, J., Pape, D., Wyllie, T., Li, L.,
 Kissinger, J., Roos, D.S., Maria, M., Hillier, L., Chinwalla, A.,
 Blistain, A., Schmitt, A., Person, B., Theising, B., Ritter, E., Ronko,
 I., Bennett, J., Cole, R., Underwood, K., Cardenas, M., Gibbons, M.,
 Harvey, N., McCann, R., Tsagaris, V., Williams, T., Jackson, Y.,
 Bowers, Y., Swaller, T., Waterston, R., Wilson, R., and Beverley, S.M.

A survey of the Leishmania major Friedlin strain VI genome by

JOURNAL
COMMENT

shotgun sequencing: a resource for DNA
Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)
Other_GSSs: lml7b03.xl
Contact: Akopyants, NS / Beverley, SM
Washu Leishmania Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Library construction: Natalia S. Akopyants, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
If using this information please cite:
N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major
Friedlin strain VI genome by shotgun sequencing' and the Washington
University Genome Sequencing Center For Information on obtaining
clone material please contact: Natalia S. Akopyants Ph.D.
(nataliab@wustl.edu) and/or Stephen M. Beverley Ph.D.
(beverley@borcim.wustl.edu)
Seq primer: -40RP from Gibco
Class: Shotgun
High quality sequence stop: 373.
Location/Qualifiers

FEATURES

source

1..461
/organism="Leishmania major"
/strain="Friedlin strain VI"
/db_xref="taxon:5664"
/clone="LMAJFV1_lml7b03"
/clone_lib="Leishmania major FV1 random genomic library"
/lab_host="TOP10 (Invitrogen)"
/note="Vector: pZero-2 (Invitrogen); Site: 1: EcoRV;
Genomic DNA was isolated from stationary phase cells. For
this library, DNA was sheared to give a tight size
distribution of 1-1.5kb fragments, blunt-ended with T4 DNA
polymerase, dephosphorylated with Shrimp Alkaline
Phosphatase and ligated into pZero-2 vector's EcoRV site."

BASE COUNT

103 a 131 c 139 g 88 t

Query Match 72.8%; Score 18.2; DB 234; Length 461;
Best Local Similarity 87.0%; Pred. No. 4.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gatcctgcaggtgcagtcacg 24
||||| ||||| ||||| |||||
Db 83 GATCTTGTCAGCTGACGTCAG 61

RESULT

6

CNS026T2 872 bp DNA GSS 12-MAY-2000
LOCUS Tetradon nigroviridis genome survey sequence pUC-ori end of clone
DEFINITION 241J15 of library G from Tetradon nigroviridis, genomic survey
sequence.

ACCESSION AL183824.1 GI:7821928
VERSION AL183824.1
KEYWORDS GSS: genome survey sequence.
SOURCE Tetradon nigroviridis.
ORGANISM Tetradon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes;
Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 872)
Rost-Crollius,H., Jalllon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.

TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigroviridis

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 872)
Rost-Crollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C.,

Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetradon nigroviridis DNA sequence

JOURNAL Unpublished
REFERENCE 3 (bases 1 to 872)
AUTHORS

TITLE Direct Submission
JOURNAL Submitted (12-Apr-2000) to the EMBL/Genbank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetradon>.

FEATURES

source

1..872
/organism="Tetradon nigroviridis"
/db_xref="taxon:99883"
/clone="241J15"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG241CE08SP1-end :
pUC-ori"

BASE COUNT

139 a 301 c 254 g 166 t 12 others

ORIGIN

Query Match 72.8%; Score 18.2; DB 220; Length 872;
Best Local Similarity 87.0%; Pred. No. 5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cgatcctgcaggtgcagtcacg 23
||||| ||||| ||||| |||||
Db 715 CGCTCCTCTGACGTCAGGTCGA 737

RESULT

7

AL556337 883 bp mRNA EST 16-FEB-2001
LOCUS AL556337 L17.NFL006.PL2 Homo sapiens cDNA clone CSODK001Y114 5
DEFINITION Prime, mRNA sequence.

ACCESSION AL556337
VERSION AL556337.1 GI:12898922
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 883)
H.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
unpublished (2001)

JOURNAL Contact: Genoscope
COMMENT Genoscope - Centre National de Sequencage
BP 101 91006 Evry cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..883
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSODK001Y114"
/clone_lib="L17.NFL006.PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site: 1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : liang@litech.com URL :
<http://fulllength.invitrogen.com>

BASE COUNT 193 a 237 c 228 g 224 t 1 others

Query Match 72.8%; Score 18.2; DB 106; Length 883;
 Best Local Similarity 87.0%; Pred. No. 5e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 gatctctgacagtgacagtcag 24
 || ||||| ||||| ||||| |||||
 Db 33 GAGCCTCGCAGGTGCGGTGAG 17

RESULT 8
 LOCUS BG023733 904 bp mRNA EST 24-JAN-2001
 DEFINITION 602303347F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4395058 5',
 mRNA sequence.
 ACCESSION BG023733
 VERSION BG023733.1 GI:12408790
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 904)
 NIH-MGC http://mgs.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LHAM10091 row: m column: 11
 High quality sequence stop: 733.
 Location/Qualifiers
 1..904
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="4395058"
 /clone_lib="NIH_MGC_88"
 /tissue_type="thudonal adenocarcinoma, cell line"
 /note="Organ: small intestine; Vector: pCMV-SPORT6;
 Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
 oligo-dt primed. Average insert size 1.767 kb. Library
 enriched for full-length clones and constructed by Life
 Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 190 a 253 c 239 g 222 t

ORIGIN

Query Match 72.8%; Score 18.2; DB 172; Length 904;
 Best Local Similarity 87.0%; Pred. No. 5e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 gatctctgacagtgacagtcag 24
 || ||||| ||||| ||||| |||||
 Db 33 GAGCCTCGCAGGTGCGGTGAG 11

RESULT 9
 LOCUS AL580709 917 bp mRNA EST 16-FEB-2001
 DEFINITION AL580709 LTR_NFL008_TC2 Homo sapiens cDNA clone CS0000081W14 5
 prime, mRNA sequence.
 ACCESSION AL580709
 VERSION AL580709.1 GI:12946994
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 917)
 NIH-MGC http://mgs.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: Miklos Palakovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shlraiki
 Toshiyuki and Ptero Carninci (RIKEN)
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LHAM10036 row: f column: 13
 High quality sequence stop: 585.
 Location/Qualifiers
 1..917
 /organism="Homo sapiens"

FEATURES
 source

BASE COUNT 204 a 246 c 233 g 229 t 5 others

ORIGIN

Query Match 72.8%; Score 18.2; DB 106; Length 917;
 Best Local Similarity 87.0%; Pred. No. 5e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 gatctctgacagtgacagtcag 24
 || ||||| ||||| ||||| |||||
 Db 59 GAGCCTCGCAGGTGCGGTGAG 37

RESULT 10
 LOCUS BF979422 1095 bp mRNA EST 23-JAN-2001
 DEFINITION 602288231F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4373772 5',
 mRNA sequence.
 ACCESSION BF979422
 VERSION BF979422.1 GI:12346637
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 1095)
 NIH-MGC http://mgs.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: Miklos Palakovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shlraiki
 Toshiyuki and Ptero Carninci (RIKEN)
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LHAM10036 row: f column: 13
 High quality sequence stop: 585.
 Location/Qualifiers
 1..1095
 /organism="Homo sapiens"

/db_xref="taxon:9606"
 /clone_image="4373772"
 /clone_lib="NIH_MGC_97"
 /lab_host="DH10B"
 /note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gcgcag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to R0T 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 248 a 314 c 343 g 188 t 2 others

ORIGIN

Query Match 72.8%; Score 18.2; DB 172; Length 1095;
 Best Local Similarity 87.0%; Pred. No. 5.1e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 atctctgcaggtgcagtcag 24
 ||||| ||||| ||||| |||||
 Db 326 GAGCCTCGCAGGTGCCGTCAG 304

RESULT 11
 A0982558 494 bp DNA GSS 29-JAN-2000
 LOCUS RPCI-23-294H19.TV RPCI-23 Mus musculus genomic clone RPCI-23-294H19
 DEFINITION / DNA sequence.
 ACCESSION A0982558
 VERSION A0982558.1 GI:6814859
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 494)
 Zhao,S., Nieman,W., Feldblum,T., Malek,J., Shatsman,S., Akinet ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.
 Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)
 Other GSS: RPCI-23-294H19.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Reseach Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tldb/bac_ends/mouse/Bac_end_intro.html
 Plate: 294 row: H column: 19
 Seq primer: SP6
 Class: BAC ends.

FEATURES
 source

Location/Qualifiers
 1..494
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-294H19"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested

with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 167 a 94 c 130 g 102 t 1 others

ORIGIN

Query Match 71.2%; Score 17.8; DB 236; Length 494;
 Best Local Similarity 90.5%; Pred. No. 7.2e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 3 atctctgcaggtgcagtcag 23
 ||||| ||||| ||||| |||||
 Db 98 ATCTCTGTCAGTTGCAGGACA 78

RESULT 12
 BE234470 536 bp mRNA EST 10-JUL-2000
 LOCUS BE234470
 DEFINITION 141586 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
 ACCESSION BE234470
 VERSION BE234470.1 GI:9019188
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 1 (bases 1 to 536)
 Fahrenkrug,S.C., Freling,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grose,W.M., Bennett,G.A., Laegfeld,W.W. and Keele,J.W.
 Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
 Unpublished (2000)
 Contact: Smith TP1
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt-trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.
 PCR primers
 FORWARD: AGGAACAGCTATGACCAAT
 BACKWARD: GTTTCACAGTCACGACG
 Plate: 79 row: F column: 18
 Seq primer: ATTATGGTGCACCTATAG.

FEATURES
 source

Location/Qualifiers
 1..536
 /organism="Sus scrofa"
 /db_xref="taxon:9623"
 /clone_lib="MARC 1P1G"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos."

BASE COUNT 92 a 126 c 170 g 146 t 2 others

ORIGIN

Query Match 71.2%; Score 17.8; DB 165; Length 536;
 Best Local Similarity 90.5%; Pred. No. 7.3e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 3 atctctgcaggtgcagtcag 23
 ||||| ||||| ||||| |||||
 Db 162 ATCTCTGTCAGTTGCAGGACA 182

RESULT 13

AL538687
 LOCUS AU538687 777 bp mRNA EST 16-FEB-2001
 DEFINITION AU538687 LTI.FL013.FBrn1 Homo sapiens cDNA clone CSDDF037P11.3
 ACCESSION AL538687
 VERSION AU538687.1 GI:12867208
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 777)
 AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 Location/Qualifiers
 1. 777
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CSDDF037P11"
 /clone_lib="LTI.FL013.FBrn1"
 /dev_stage="pooled tissue from post conception fetuses (20 week, 24 week and 26 week)"
 /lab_host="DH10B"
 /note="Organ: Fetal brain; Vector: PCWSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dt) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the PCWSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville , Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
 BASE COUNT 173 a 220 c 205 g 169 t 10 others
 ORIGIN
 Query Match 71.2%; Score 17.8; DB 106; Length 777;
 Best Local Similarity 82.6%; Pred. No. 7.5e+02;
 Matches 19; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 2 gatcctgcagtcagtcagtcag 24
 |||||
 Db 486 GGTCTCTGTCAGACAGTCAG 508
 |||||
 RESULT 14
 LOCUS CNS044U0 1032 bp DNA GSS 18-MAY-2000
 DEFINITION Tetradon nigroviridis genome survey sequence T7 end of clone 082E16 of library G from Tetradon nigroviridis, genomic survey sequence.
 ACCESSION AL274571
 VERSION AL274571.1 GI:8008771
 KEYWORDS GSS; genome survey sequence.
 SOURCE Tetradon nigroviridis.
 ORGANISM Tetradon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetradon.
 REFERENCE 1 (bases 1 to 1032)
 AUTHORS Roest-Crollius, H., Jallou, O., Dasilva, C., Fitzames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.
 TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis
 JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1032)
 AUTHORS Roest-Crollius, H., Jallou, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fitzames, C., Wincker, P., Brotier, P., Quetier, F., Saurin, W. and Weissenbach, J.
 TITLE Human gene number estimate provided by genome wide analysis using Tetradon nigroviridis DNA sequence
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 1032)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
 COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetradon.
 Location/Qualifiers
 1. 1032
 /organism="Tetradon nigroviridis"
 /db_xref="taxon:99883"
 /clone="082E16"
 /clone_lib="G"
 /note="Genoscope sequence ID : COBG082BC08LP1-end : T7"
 BASE COUNT 267 a 249 c 220 g 261 t 35 others
 ORIGIN
 Query Match 71.2%; Score 17.8; DB 221; Length 1032;
 Best Local Similarity 90.5%; Pred. No. 7.6e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 4 tctcgtcagtcagtcagtcag 24
 |||||
 Db 753 TCCTCAGCAGTCAGTCAG 773
 |||||
 RESULT 15
 LOCUS AA594461 339 bp mRNA EST 25-SEP-1997
 DEFINITION N194C07.S1 NCI_CGAP Co10 Homo sapiens cDNA clone IMAGE:1058316 3' similar to gb:M77198 RAC-BETA SERINE/THREONINE KINASE (HUMAN),, mRNA sequence.
 ACCESSION AA594461
 VERSION AA594461.1 GI:2409811
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 339)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgap@bbs-riemail.nih.gov
 Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bdpr/image/image.html
 Trace considered overall poor quality
 Insert Length: 2573 Std Error: 0.00
 Seq Primer: 40m13 fwd. ET from Amersham
 High quality sequence stop: 1.
 Location/Qualifiers
 1. 339
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1058316"

```

/clone_lib="NCI CGAP.C010"
/tissue_type="colon tumor RER+"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
RER+ colon tumor, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo (N-Soares4)."

BASE COUNT      99 a      76 c      96 g      68 t
ORIGIN

Query Match      70.4%; Score 17.6; DB 9; Length 339;
Best Local Similarity 83.3%; Pred. No. 8.6e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2      gatcctcgtcaggtcagtcagc 25
        ||||| ||||| | ||| |||||
Db      45      GATCGCGTCACGCGCAGCTCAGC 68

RESULT 16
BF318689      360 bp      mRNA      EST      29-DEC-2000
LOCUS      BF318689/c
DEFINITION      uy05a01.y1 McCarrey Eddy spermatoocytes Mus musculus cDNA clone
IMAGE:3657096 5' similar to TR:Q9VB28 Q9VB28 CGI3972 PROTEIN. ;
RNA sequence.
ACCESSION      BF318689
VERSION      BF318689.1 GI:11267302
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
1 (bases 1 to 360)
Marrin, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person
, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Riltter
, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Maria M/Mashu-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:149400
High quality sequence stop: 338.

FEATURES
Source
Location/Qualifiers
1..360
/organism="Mus musculus"
/strain="CD-1"
/db_xref="taxon:10090"
/clone_image:3657096"
/clone_lib="McCarrey Eddy spermatoocytes"
/sex="male"
/tissue_type="spermatoocytes, pooled from multiple mice"
/dev_stage="60 day"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pBluescript SK+ (Stratagene
); Site.1: XhoI; Site.2: EcoRI. cDNA oligo dT-primed
[5'-(GA)10-ACTAGTCGACGTTTCTTTTCTTTT-3'] and directionally
cloned using 5' linkers 5'-AATTGCGCAGCAG-3' and
5'-CTGTCGCGC-3'. Size selection of >400bp material gives
average insert size ranging from 1-2 kb. Library was mass
excised (from lambda-Unizap-XR) and resulting

```

```

single-stranded phagemids were prepped and tranformed
into DH10B. Library contains 984 recombinants.
References: J. Androl. 20:635-639 and Gene 25:263-269.
Library constructed and donated by J. McCarrey, Ph.D.
(Southwest Foundation for Biomedical Research, Dept. of
Genetics); excision done by E.M. Eddy, Ph.D. (National
Institutes of Health, National Institute of Environmental
Health Sciences). Original lambda-based library is
available through ATCC, catalog #63422."

BASE COUNT      118 a      71 c      115 g      55 t
ORIGIN

Query Match      70.4%; Score 17.6; DB 147; Length 360;
Best Local Similarity 80.0%; Pred. No. 8.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1      cgatcctcgtcaggtcagtcagc 25
        ||||| ||||| | ||| |||||
Db      181      CGATCCTCCTCAGCTNCAGGTCATC 157

RESULT 17
BF659032      369 bp      mRNA      EST      20-DEC-2000
LOCUS      maa89h11.y1 Soares thymus.2NBMT Mus musculus cDNA clone
DEFINITION      IMAGE:3824132 5' similar to SW:PIK_HUMAN P42356
PHOSPHATIDYLINOSITOL 4-KINASE ALPHA ;, mRNA sequence.
ACCESSION      BF659032
VERSION      BF659032.1 GI:11924166
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 369)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: ccapbs_r@mail.nih.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1460484
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 347.

FEATURES
Source
Location/Qualifiers
1..369
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image:3824132"
/clone_lib="Soares_thymus_2NBMT"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAGTGGAGCGCGCGCTTTTCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT      70 a      121 c      104 g      74 t
ORIGIN

```


20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: MEST@mail.nih.gov
 cDNA Library Preparation: M.B. Soares Lab Clone distribution:
 Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
 should be noted that Bento Soares is generating a small number of
 additional specialized non-redundant arrays of BMAP cDNAs whose
 availability will be considered under appropriate and limited
 collaborative arrangements
 Seq primer: M13 Reverse.
 Location/Qualifiers

FEATURES

source
 1..413
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U1-M-BH0-ajf-b-11-0-U1"
 /clone_lib="NIH-BMAP_M_S1"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /note="vector: pTR73D-Pac (Pharmacia) with a modified
 polylinker: Site 1: Not I; Site 2: Eco RI; The
 NIH-BMAP_M_S1 library is a subtracted library derived from
 a mixture of normalized libraries from ten regions of the
 mouse brain (cerebellum, brain stems, olfactory bulbs,
 hypothalamus, cortex, amygdala, basal ganglia, pineal
 gland, striatum, hippocampus). The driver used for
 subtraction consisted of a pool of 20,000 cDNA clones
 obtained from non-normalized and normalized libraries of
 these ten regions of the mouse brain."
 BASE COUNT 92 a 125 c 112 g 84 t
 ORIGIN

Query Match 70.4%; Score 17.6; DB 141; Length 413;
 Best Local Similarity 83.3%; Pred. No. 8.8e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 cgatctctgacagtcagtcagtcag 24
 ||||| ||||| ||||| ||||| |||||
 Db 24 CAATGTCGTCAGTGCAGTGCAG 1

RESULT 21
 LOCUS A0172663 424 bp DNA GSS 17-OCT-1998
 DEFINITION HS_3194_A2_D09_MR CIT Approved Human Genomic Sperm Library D Homo
 sapiens genomic clone Plate=3194 Col=18 Row=G, DNA sequence.
 ACCESSION A0172663
 VERSION A0172663.1 GI:3570030
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 424)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 Title: Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 3194 Row: G Column: 18
 Class: BAC ends

High quality sequence stop: 424.
 Location/Qualifiers
 1..424
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=3194 Col=18 Row=G"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelBAC11; BAC Clones in
 E-coli DH10B"

BASE COUNT 119 a 110 c 94 g 101 t
 ORIGIN

Query Match 70.4%; Score 17.6; DB 225; Length 424;
 Best Local Similarity 83.3%; Pred. No. 8.8e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 cgatctctgacagtcagtcagtcag 24
 ||||| ||||| ||||| ||||| |||||
 Db 382 CAATGTCGTCAGTGCAGTGCAG 359

RESULT 22
 LOCUS A0371411/c 429 bp DNA GSS 06-MAR-1999
 DEFINITION HS_5043_A1_D05_SP66 RPII1 Human Male BAC Library Homo sapiens
 genomic clone Plate=619 Col=9 Row=G, DNA sequence.
 ACCESSION A0371411
 VERSION A0371411.1 GI:4340516
 KEYWORDS GSS.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 429)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 Title: Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end web server: http://www.hsc.washington.edu
 Plate: 619 Row: G Column: 9
 Seq primer: SP6
 Class: BAC ends
 High quality sequence stop: 429.

FEATURES

source
 1..429
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=619 Col=9 Row=G"
 /clone_lib="RPII1 Human Male BAC Library"
 /sex="male"
 /cell_type="Lymphocytes"
 /note="vector: pBAC63.6; RPII1 Human Male BAC Library"
 BASE COUNT 122 a 90 c 109 g 107 t
 ORIGIN

Query Match 70.4%; Score 17.6; DB 227; Length 429;
 Best Local Similarity 83.3%; Pred. No. 8.8e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

REFERENCE	AUTHORS	TITLE	JOURNAL
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	1 (bases 1 to 618)		
	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Gessel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.	The WashU-HMF Mouse EST Project	Unpublished (1996)

Email: mesfr@mail.nih.gov
CDNA Library Preparation: M.B. Soares lab Clone distribution:
Researchers may obtain BMP cDNA clones from RESEARCH GENETICS. It
should be noted that Beto Soares is generating a small number of
additional specialized non-redundant arrays of BMP cDNAs whose
availability will be considered under appropriate and limited
collaborative arrangements

RESULT 35
CNS006PE/C 982 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC # BACR14J05 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL065656
VERSION AL065656.1 GI:4944725
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 982)
Genoscope.
AUTHORS Direct Submission
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
JOURNAL - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoler in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1. 982
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR14J05"
/note="end : 77"
BASE COUNT 205 a 289 c 287 g 161 t 40 others
ORIGIN

Query Match 70.4% Score 17.6; DB 219; Length 982;
Best Local Similarity 83.3%; Pred. No. 9.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 gatctcgtcaggtgcaggtcagc 25
||||| ||||| || ||||| |||||
Db 116 GATCCAGCTCGGAGCGGCTCAGC 93

RESULT 36
CNS00HWP/C 985 bp DNA GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR35J21 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL073836
VERSION AL073836.1 GI:4953810
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 985)
Genoscope.
AUTHORS Direct Submission
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

COMMENT
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoler in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1. 985
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR35J21"
/note="end : TET3"
BASE COUNT 204 a 275 c 300 g 164 t 42 others
ORIGIN

Query Match 70.4% Score 17.6; DB 219; Length 985;
Best Local Similarity 83.3%; Pred. No. 9.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 gatctcgtcaggtgcaggtcagc 25
||||| ||||| || ||||| |||||
Db 117 GATCCAGCTCGGAGCGGCTCAGC 94

RESULT 37
BF965985/C 1002 bp mRNA EST 22-JAN-2001
LOCUS 602277168F2 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4364968 5',
DEFINITION mRNA sequence.
ACCESSION BF965985
VERSION BF965985.1 GI:12333200
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1002)
NIH-MGC <http://mgs.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM10013 row: g column: 17
High quality sequence stop: 570.
Location/Qualifiers
1. 1002
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4364968"
/clone_lib="NIH_MGC_86"
/clone="BF965985"
/tissue_type="osteosarcoma, cell line"
/lab_post="DH10B (phage-resistant)"
/note="Organ: bone; Vector: pCMV-SPORT6; Site: 1; NotI;

FEATURES
source

Site-2: Sail; Cloned unidirectionally; oligo-dr primed.
Average insert size 1.533 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC library."

BASE COUNT 266 a 266 c 288 g 182 t

Query Match 70.4%; Score 17.6; DB 171; Length 1002;
Best Local Similarity 83.3%; Pred. No. 9.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 gatctcgtcagtgccagtcagc 25
||||| ||||| ||||| ||||| |||||
Db 649 GCTCCTCTCAGAGCAGGCGCAGC 626

RESULT 38
LOCUS CNS00H01 1016 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC:
BACR35124 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL074118
AL074118.1 GI:4953595
GSS.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1016)
Genoscope.

AUTHORS
TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seque@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazuo Osoegawa and
Aaron Mamoser in Pletier de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
PI and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
SOURCE
Location/Qualifiers
1..1016
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR35124"
/note="end : TET3"

BASE COUNT 201 a 295 c 299 g 169 t 52 others

Query Match 70.4%; Score 17.6; DB 219; Length 1016;
Best Local Similarity 83.3%; Pred. No. 9.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 gatctcgtcagtgccagtcagc 25
||||| ||||| ||||| ||||| |||||
Db 118 GATCAGCTCGGAGCGGCTCAGC 95

RESULT 39

BE795069/c
LOCUS BE795069 1126 bp mRNA EST 20-SEP-2000
DEFINITION 601592885F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3946869 5',
mRNA sequence.
BE795069
ACCESSION BE795069.1 GI:10216267
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1126)
NIH-MGC <http://imgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LICM807 row: j column: 22
High quality sequence start: 10
High quality sequence stop: 702.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES
SOURCE
Location/Qualifiers
1..1126
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3946869"
/clone_lib="NIH_MGC_7"
/tissue="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site: 1: XhoI, Site-2:
EcoRI; CDNA made by oligo-dr priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 262 a 320 c 289 g 255 t

Query Match 69.6%; Score 17.4; DB 140; Length 1126;
Best Local Similarity 94.7%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gatctcgtcagtgccagtcagc 20
||||| ||||| ||||| ||||| |||||
Db 853 GATCCTCTCAGAGCAGGCTCAGC 835

RESULT 40
LOCUS BG328084 1298 bp mRNA EST 27-FEB-2001
DEFINITION 602427134F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4546497 5',
mRNA sequence.
BG328084
ACCESSION BG328084.1 GI:13134522
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1298)
NIH-MGC <http://imgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Email: c9apbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: NIH Intramural Sequencing Center
 Clone distribution: MCC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LICM123 row: c column: 10
 High quality sequence stop: 511.

FEATURES

Source

Location/Qualifiers
 1..1298
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4546497"
 /clone_lib="NIH_MGC_15"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pOTB7; Site:1; XhoI; Site:2;
 EcoRI; CDNA made by oligo-dT priming, directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGACGAG(G). Size selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-CDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)"
 BASE COUNT 316 a 498 c 357 g 127 t
 ORIGIN

Query Match 69.8%; Score 17.4; DB 152; Length 1298;
 Best Local Similarity 94.7%; Pred. No. 1.2e+03;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 4 tctctgcagtcagtcagtc 22
 ||| ||||| ||||| |||||
 Db 987 TCGTCGTCAGGTCGAGTC 969

RESULT 41

AM371554 108 bp mRNA EST 04-FEB-2000
 LOCUS
 DEFINITION QY3-BT0295-111199-017-g08 BT0295 Homo sapiens cDNA, mRNA sequence.
 ACCESSION AM371554
 VERSION
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 108)
 AUTHORS HCCP <http://www.ludwig.org.br/ORESTES>.
 TITLE The FAPESP/LICR Human Cancer Genome Project
 JOURNAL Unpublished (1999)
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil

Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?l-QY3&t2-QY3-BT0295-111199-017-g08&t3-1999-11-11&t4-1>)
 Seq primer: puc 18 forward
 High quality sequence stop: 108.

FEATURES

Source

1..108
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="BT0295"

BASE COUNT 8 a 39 c 23 g 38 t
 ORIGIN

Query Match 68.8%; Score 17.2; DB 115; Length 108;
 Best Local Similarity 86.4%; Pred. No. 1.2e+03;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 tctctgcagtcagtcagtc 25
 ||||| ||||| ||||| |||||
 Db 1 TCCTCCTCAGACGACGACG 22

RESULT 42
 AA651689 112 bp mRNA EST 13-NOV-1997
 LOCUS
 DEFINITION nm47b06.r1 NCL-CGAP_Kid6 Homo sapiens cDNA clone IMAGE:108695 5'
 similar to gb:X63526 ELONGATION FACTOR 1-GAMMA (HUMAN); mRNA
 sequence.

ACCESSION AA651689
 VERSION
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 112)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-remail.nih.gov
 Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Stratagene, Inc.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 815 Std Error: 0.00
 Seq primer: -28m13 rev1 ET from Amersham.

FEATURES

Source

Location/Qualifiers
 1..112
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:108695"
 /clone_lib="NCI-CGAP_Kid6"
 /sex="mixed"
 /tissue_type="Kidney tumor"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: kidney; Vector: Bluescript SK-; Site:1;
 EcoRI; Site:2; XhoI; Cloned unidirectionally. Primer:
 oligo dT. Pooled kidney tumors, 5' adaptor sequence: 5'
 GAATTCGGACGAG 3' 3' adaptor sequence: 5'
 CTCGAGTCTTTTTTTTTTTTTTT 3' Average insert size: 1.0 kb."
 BASE COUNT 44 a 21 c 36 g 11 t
 ORIGIN

Query Match 68.8%; Score 17.2; DB 10; Length 112;
 Best Local Similarity 86.4%; Pred. No. 1.2e+03;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 tctctgacagtgacagtcagc 25
 ||||| ||||| ||||| |||||
 Db 101 TCCTCTCAGGACGACGACG 80

RESULT 43
 LOCUS AAI97041/c
 DEFINITION zq10c11.1 Stratiagene muscle 937209 Homo sapiens cDNA clone
 IMAGE:629300 5' similar to gb:X63526 ELONGATION FACTOR 1-GAMMA
 (HUMAN); mRNA sequence.

ACCESSION AAI97041
 VERSION AAI97041.1 GI:1792665
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 235)
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
 Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
 White,Y., Wylie,T., Waterston,R. and Wilson,R.
 TITLE WashU-NCI human EST project
 JOURNAL Unpublished (1997)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 757 Std Error: 0.00
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 138.

FEATURES
 Source Location/Qualifiers
 1..235
 /organism="Homo sapiens"
 /db_xref="GDB:5049534"
 /db_xref="taxon:9606"
 /clone="IMAGE:629300"
 /clone_lib="Stratiagene muscle 937209"
 /tissue_type="muscle"
 /dev_stage="adult"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: skeletal muscle; Vector: pBluescript SK-;
 Site:1: EcoRI, Site:2: XhoI; Cloned unidirectionally.
 Primer: Oligo dT. Skeletal muscle from patient with
 malignant hyperthermia. Average insert size: 1.0 kb;
 Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGCGCAGCAG
 3'-3' adaptor sequence: 5' CTCGACTTTTCTTTTCTTTT 3'"

BASE COUNT 56 a 52 c 71 g 45 t 11 others
 ORIGIN

Query Match 68.8%; Score 17.2; DB 3; Length 235;
 Best Local Similarity 86.4%; Pred. No. 1.3e+03;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 tctctgacagtgacagtcagc 25
 ||||| ||||| ||||| |||||
 Db 83 TCCTCTCAGGACGACGACG 62

RESULT 44
 LOCUS BF349456
 DEFINITION RC2-HT0148-021099-012-g03 HT0148 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF349456
 VERSION BF349456.1 GI:11308530
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 236)
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Britones,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.V. and
 Simpson,A.J.
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=RC2&t2=RC2-HT0148-021099-012-g03&t3=1999-10-02&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence stop: 8.

FEATURES
 Source Location/Qualifiers
 1..236
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="HT0148"
 /clone="RC2-HT0148"
 /dev_stage="Adult"
 /note="Organ: head,neck; Vector: puc18; Site:1: SmaI;
 Site:2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

BASE COUNT 68 a 40 c 58 g 70 t
 ORIGIN

Query Match 68.8%; Score 17.2; DB 147; Length 236;
 Best Local Similarity 86.4%; Pred. No. 1.3e+03;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 tctctgacagtgacagtcagc 25
 ||||| ||||| ||||| |||||
 Db 47 TCCTCTCAGGACGACGACG 68

RESULT 45
 LOCUS NB3353/c
 DEFINITION EC0312F Human fetal heart, lambda ZAP Express Homo sapiens cDNA
 clone EC0312 5' similar to ELONGATION FACTOR 1-GAMMA, mRNA
 sequence.

ACCESSION NB3353
 VERSION NB3353.1 GI:1258978
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

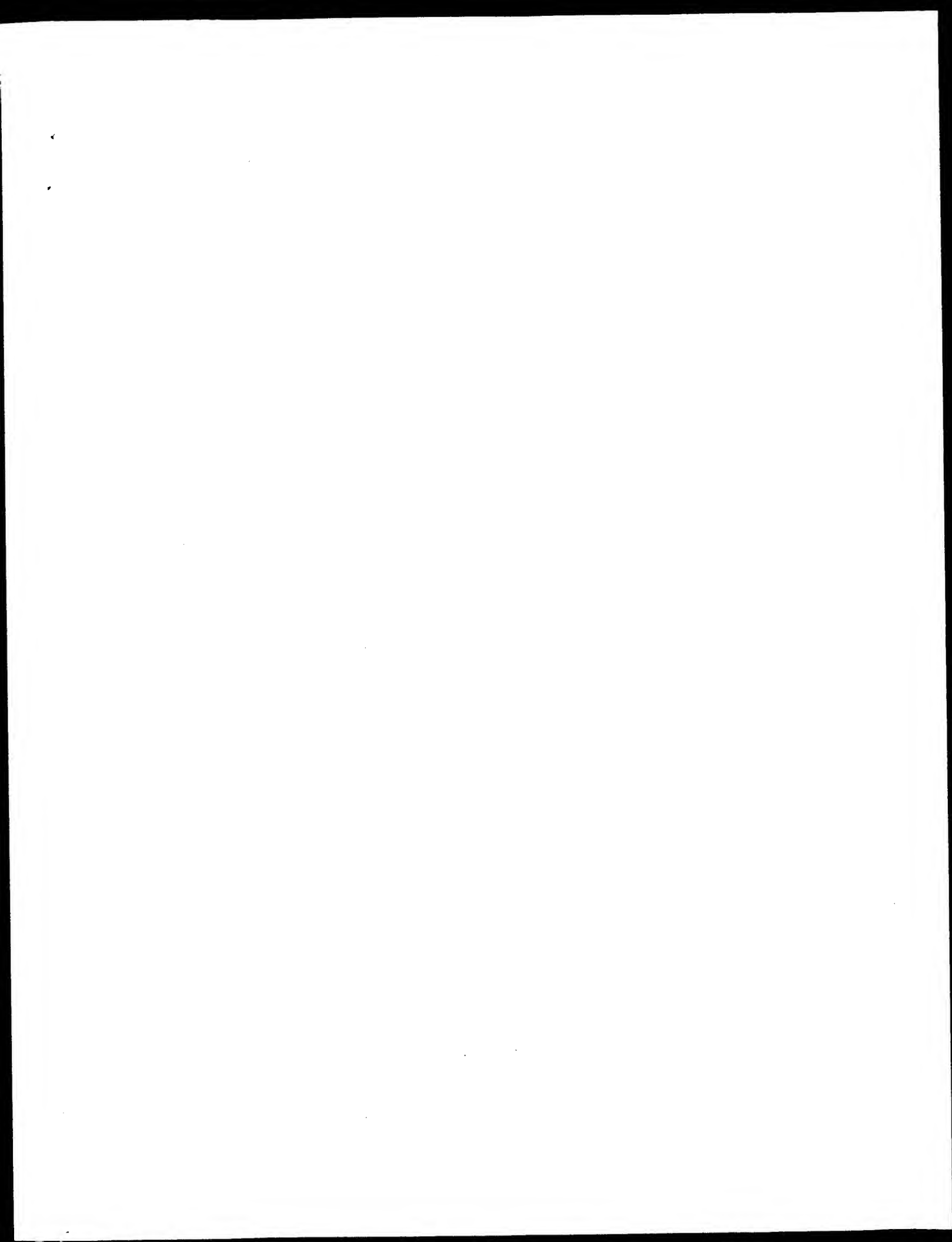
REFERENCE 1 (bases 1 to 255)
 AUTHORS Liew,C.C.
 TITLE cDNAs from fetal heart (1996)
 JOURNAL Unpublished (1996)
 COMMENT Contact: Liew CC
 Department of Laboratory Medicine and Pathobiology

University of Toronto
 Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
 Tel: 4169788758
 Fax: 4169785650
 Email: liewce@utcc.utoronto.ca
 Seq primer: GAATTTAACCTCAGCTAAAGG.
 Location/Qualifiers
 1..255
 source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="EC0312"
 /lab_host="E. coli XL1-Blue"
 /note="Vector: lambda ZAP Express; Site_1: EcoRI; Site_2:
 XhoI; mRNA was purified from human fetal hearts (8-10
 weeks). cDNA was synthesized using a XhoI-0190 dt
 adaptor-primer. EcoRI adaptors were ligated, followed by
 digestion with XhoI, for directional cloning into
 predigested lambda ZAP Express."
 BASE COUNT 76 a 58 c 80 g 41 t
 ORIGIN

Query Match 68.8%; Score 17.2; DB 159; Length 255;
 Best Local Similarity 86.4%; Pred. No. 1.3e+03;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 0y 4 tccctcaggtgcaggtcagc 25
 ||||| ||||| ||||| |||||
 Db 229 TCCTCTCAGGAGCAGGCGCAGC 208

Search completed: October 9, 2001, 13:46:52
 Job time: 9527 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:40:14 ; Search time 1670.83 Seconds
(Without alignments)
231.438 Million cell updates/sec

Title: US-09-396-196f-1

Perfect score: 25

Sequence: 1 gacattgcgaagtcacagaattta 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 9

Total number of hits satisfying chosen parameters: 188662

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_cm:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rnd:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rnd:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_cm:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vl:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vl1:*
59: gb_vl2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_pr10:*
95: gb_pr11:*
96: gb_pr12:*
97: gb_pr13:*
98: em_ba3:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	25	100.0	1041	9 AR029499	AR029499 Sequence
2	25	100.0	1041	9 AR034916	AR034916 Sequence
3	25	100.0	1084	9 A11530	A11530 B10B gene o
4	25	100.0	1121	10 E00893	E00893 Genomic DNA
5	25	100.0	5793	2 E00893	J04423 E.coli 7,8-
6	25	100.0	5872	9 A38246	A38246 Sequence 1
7	25	100.0	5872	9 A38251	A38251 Sequence 6
8	25	100.0	5872	9 A93674	A93674 Sequence 1

```

9      25      100.0      5872      9      A93679      Sequence 6
10     25      100.0      5872      9      AR101809      Sequence
11     25      100.0      5872      9      AR101810      Sequence
12     25      100.0      11022      1      AE000180      Escherich
13     25      100.0      13501      1      AE005258      Escherich
14     25      100.0      297816      2      AP000253      Escherich
15     21      84.0      5526      2      AF250776      Unculture
16     18      72.0      8227      2      AF248314      Unculture
17     16      64.0      32767      77      AC087305      Homo sapi
18     16      64.0      155172      66      AC021110      Homo sapi
19     16      64.0      156184      71      AC041018      Homo sapi
20     16      64.0      174815      82      AP001390      Homo sapi
21     16      64.0      196172      66      AC020734      Homo sapi
22     15      60.0      1771      9      AR080743      Sequence
23     15      60.0      1771      9      AR080743      Sequence
24     15      60.0      1829      54      M0SINTESUB      L20788 Mus musculu
25     15      60.0      1949      94      MMINTEGRO1      U34627 Mus musculu
26     15      60.0      2340      2      ECHK31TM      X82231 E.coli HK31
27     15      60.0      3748      10      AX098224      Sequence
28     15      60.0      4602      91      D13644      Human mRNA
29     15      60.0      4636      1      AF039582      AF039582 Enterobac
30     15      60.0      5218      89      AF152313      AF152313 Homo sapi
31     15      60.0      5221      89      AF152315      AF152315 Homo sapi
32     15      60.0      5233      89      AF152308      AF152308 Homo sapi
33     15      60.0      5251      89      AF152312      AF152312 Homo sapi
34     15      60.0      5254      89      AF152306      AF152306 Homo sapi
35     15      60.0      5254      89      AF152310      AF152310 Homo sapi
36     15      60.0      5257      89      AF152307      AF152307 Homo sapi
37     15      60.0      5260      89      AF152305      AF152305 Homo sapi
38     15      60.0      5260      89      AF152309      AF152309 Homo sapi
39     15      60.0      5260      89      AF152311      AF152311 Homo sapi
40     15      60.0      5260      89      AF152314      AF152314 Homo sapi
41     15      60.0      5260      89      AF152316      AF152316 Homo sapi
42     15      60.0      5260      89      AF152317      AF152317 Homo sapi
43     15      60.0      5299      89      AF152303      AF152303 Homo sapi
44     15      60.0      5431      89      AF152304      AF152304 Homo sapi
45     15      60.0      5562      85      AB011138      Homo sapi

```

ALIGNMENTS

```

RESULT 1
LOCUS   AR029499      1041 bp      DNA
DEFINITION   Sequence 7 from patent US 5859335.
ACCESSION   AR029499
VERSION     AR029499.1      GI:5941472
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 1041)
AUTHORS     Patton,D.Andrew.
TITLE       Enhanced biotin biosynthesis in plant tissue
JOURNAL     Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
     source             Location/Qualifiers
     ..1041
     /organism="unknown"
BASE COUNT      262 a      273 c      305 g      201 t
ORIGIN

```

```

Query Match      100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 gacattgtcgaagtcacagaatta 25
        |||||||||||||||||||||||
Db      21 GACATTGTGCGAAGTCACAGAATTA 45

```

RESULT 2

```

AR034916      AR034916      1041 bp      DNA
LOCUS
DEFINITION   Sequence 7 from patent US 5869719.
ACCESSION   AR034916
VERSION     AR034916.1      GI:5950521
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 1041)
AUTHORS     Patton,D.A.
TITLE       Transgenic plants having increased biotin content
JOURNAL     Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
     source             Location/Qualifiers
     ..1041
     /organism="unknown"
BASE COUNT      262 a      273 c      305 g      201 t
ORIGIN

```

```

Query Match      100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 gacattgtcgaagtcacagaatta 25
        |||||||||||||||||||||||
Db      21 GACATTGTGCGAAGTCACAGAATTA 45

```

```

RESULT 3
LOCUS   A11530      1084 bp      DNA
DEFINITION   Biot gene of E.coli with primers.
ACCESSION   A11530
VERSION     A11530.1      GI:490218
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM    Escherichia coli.
REFERENCE   1 (bases 1 to 1084)
AUTHORS
JOURNAL
FEATURES
     source             Patent: GB 2216530-A 16 11-OCT-1989;
     ..1084
     Location/Qualifiers
     ..1084
     /organism="Escherichia coli"
     /db_xref="taxon:562"
     24..1064
     /gene="BIOB"
     24..1064
     /gene="BIOB"
     /codon_start=1
     /transl_table=1
     /protein_id="CAA00965.1"
     /db_xref="GI:490219"
     /db_xref="SMISS-PROT:P12996"
     /translation="MAHRPRTVSOVELFEKKPLDLIFPAQOVYHQRHPPROYST
     LLSITGACPECDKCYCPQSSRYRTGLEAEKLMYBOYLESARRAKAGSTRFCGAW
     KNHFERDMPFLBQWVGKAMGJEACWTJLGLTESQAOURLANGLDYNNHNTDTPSPF
     VGNITTRTYOERLDLEKVRDAGIKVSGGIVGLGVTVDRAGLLIOLIANLPTPES
     VPIIMLVKVGTPPLADNDVDADFETITVAVRIMMPTSVRLSAGREQWNEQTOAMC
     FMAGANSIFYGCKILITTPNEPEDKDLQLFKRLGLINPQGFVILAGNDQQRRLQALMT
     PTFDEYVNAAL"

```

```

BASE COUNT      271 a      286 c      318 g      209 t
ORIGIN

```

```

Query Match      100.0%; Score 25; DB 9; Length 1084;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 gacattgtcgaagtcacagaatta 25
        |||||||||||||||||||||||

```

Db 44 GACATTGTCCGCAAGTCACAGATTTA 68

RESULT 4

LOCUS E00893 1121 bp DNA PAT 29-SEP-1997

DEFINITION Genomic DNA encoding biotin Synthetase.

ACCESSION E00893

VERSION E00893.1 GI:2169154

KEYWORDS JP 1986149091-A/1.

SOURCE Escherichia coli.

ORGANISM Escherichia coli

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 1121)

AUTHORS Hirose, Y., Kojima, T. and Kimura, H.

TITLE DMPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND PRODUCTION OF BIOTIN

JOURNAL Patent: JP 1986149091-A 1 07-JUL-1986;

COMMENT NIPPON SODA CO LTD

OS Escherichia coli

PN JP 1986149091-A/1

PD 07-JUL-1986

PF 24-DEC-1984 JP 1984272605

PI HIROSE YOSHIIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC

C12N1/00, C12N1/20, C12P13/18, C12N1/20, C12R1:19), (C12P13/18, PC

C12R1:19);

CC strandedness: Double;

CC topology: Linear;

CC hypothetical: No;

CC anti-sense: No;

CC *source: strain=Escherichia coli Nsl01;

CC feature is identified by experimental;

CC Key Location/Qualifiers

FT CDS 42..1079

FT /product='biotin synthetase'

FEATURES

source Location/Qualifiers

1..1121

/organism="Escherichia coli"

/db_xref="taxon:562"

BASE COUNT 289 a 296 c 325 g 211 t

ORIGIN

Query Match 100.0%; Score 25; DB 10; Length 1121;

Best Local Similarity 100.0%; Pred. No. 1.5e-05;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gacattgtccgaagtcacagaattta 25

|||||

Db 62 GACATTGTCCGCAAGTCACAGATTTA 86

RESULT 5

ECOBIO 5793 bp DNA BCT 28-FEB-1994

LOCUS E00893

DEFINITION E.coli 7,8-diamino-pelargonic acid (bioa), biotin synthetase (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioc protein, and dehydrobiotin synthetase (biobD), complete cds.

ACCESSION J04423

VERSION J04423.1 GI:145426

KEYWORDS 7,8-diamino-pelargonic acid aminotransferase; 7-keto-8-amino-pelargonic acid synthetase; bioa gene; bioB gene; bioc gene; biobD gene; bioF gene; biotin synthetase; dehydrobiotin synthetase.

SOURCE Escherichia coli (strain K-12) DNA.

ORGANISM Escherichia coli

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 5793)

AUTHORS Otsuka, A.J., Buonocristiani, M.R., Howard, P.K., Flamm, J. and Johnson, O.

TITLE The Escherichia coli biotin biosynthetic enzyme sequences predicted

JOURNAL J. Biol. Chem. 263, 19577-19585 (1988)

MEDLINE 89066784

COMMENT Draft entry and computer-readable sequence [1] kindly submitted by A.Otsuka, 09-NOV-1988.

FEATURES

source Location/Qualifiers

1..5793

/organism="Escherichia coli"

/strain="K-12"

/db_xref="taxon:562"

/complement(98..574)

/note="ORF 1"

/codon_start=1

/transl_table=1

/protein_id="AA23513.1"

/db_xref="GI:45168"

/translation="MKLISNDLRDGDLPKRRHVFNGMGVGDGDNISPHLAWDDVAGTK SEVVCYDDPDPGSGMMHVVVNLPAADTRVLPDGFSGGLVAMPDGVLOTPTDQKGTG YDGAAPPKGETHRVLEFVHALDIERIDVDEGASAMVGFNVHFSIASITAMFS" complement(633..1925)

/gene="bioa"

/complement(633..1925)

/gene="biob"

/complement(633..1925)

/gene="bioc"

/complement(633..1925)

/transl_table=1

/product="7,8-diamino-pelargonic acid aminotransferase"

/protein_id="AA23514.1"

/db_xref="GI:45166"

/translation="MTTDLAPFDQRIHWPYTSMTSPLEVPVVSABCECLILSDGR LVDGMSWMAAIIHGVNHNPQALAMKSOIDAMSHVNGGITPAIEICRLVAMSGRN ALECVPLADSGVAIVEYKPMALQYWAQGEARORFLFRNGVHGDTCGAMVCDPDN SMHSLMKGYLPENLEFAPAROSRMDGEMDRDVGRLMAARHEIAVLEPIYOGA GGRMWHHPMLKRIKRICDREGLIILADIELATGPRCTGFCHEAELIAPOLIGKA LRGVMTLSATLITREVALETISNGEAGCMGEPPTMGNPLACAAANASLATESGDMO OGVADELVQLRDLAPARDAEMVADVRLVAGVETHPVMAALQKFFVEQGWIR PEGLIYLMPPYIILPQQLRLTAANVAVODETEFCQ"

2012..3052

/gene="biob"

2012..3052

/gene="biob"

/codon_start=1

/transl_table=1

/product="biotin synthetase"

/db_xref="GI:145425"

/translation="MAHREPWLSQVTELFEPKRLDLEFAQOYVHRQHPDROVOYT LLSIKTGACPEDECKYCPQTSRYKKTGLEARLMEVEVLSARAKAAGSTRCMGAM KNPHERMDPYLEQMOVGVRAMGLEACMTLGTLSQOQRLANAGLDYVNHNDTSPEF YGNLITTRTYOERLDITLEKVRDAGIKVCSGGIVLGTEYKVRAGLLDLQANLPTPES VPINMLYKVGKPLADNDVDADDFIRITIAVARIMPTSYVRLSAGREQMNQOTQAMC EMAGANSIFPGCKLITTPNPEEDKQDLFRKGLINPQAVIAGDNEQQRLDQLMT PVDTEYNAAL"

3049..4203

/gene="bioc"

3049..4203

/gene="bioc"

/codon_start=1

/transl_table=1

/product="7-keto-8-amino pelargonic acid synthetase"

/protein_id="AA23516.1"

/db_xref="GI:145426"

/translation="MSWQEKINAALDARRAALARRRYPAVAGAGRMVLVADDPYOLNF SNDYVGLSHHPQIIRAMOGAGAPGFGSGSGSVAGSVYHOLAEELAEWIGYSRA LPTISGPAANQOVIYAAMMAKEDRIADRSASILPAASISPOLRRFAHNDVTHLAR SCWLQKVKPELVVTEGKGFVSGAANVCSSTVADYLLQFARHLITSTMPRAQAL RASLAVIRSDGDAAREKLAALITRRAGVQDLPFLADSCSAIQPLVIGDNSRALQ L AERLROGCVWTAIRPTVPAGTARLRLTLTAHEMODIDRLLEVLHGNG"

4190..4945

/gene="bioc"

4190..4945

/gene="bioc"

gene

CDS

```

gene
CDS
/codon_start=1
/transl_table=11
/protein_id="AAA23517.1"
/db_xref="GI:145427"
/translation="MATVVKOALIAAFGRAAHYFOHADLOROSADALLAMLPORRYT
HYLDAGCGPGMSRHRERHNOVATLSPMLVOAKRQDAADHTLADISLPLATA
TPEDLMSNLAVOMCGNSTALRELYRVPRGVAFITLVGSLPERHOAVADERP
HANRFLPDELEOSLNGVYOHIOPTLWFDLISAMRSIKIGATHLHGRDPRIL
TRSOQLRLQLAMPQOQGRYPLTYHLFLGVARE"
4938..5597
/gene="biod"
/feature="biod"
4938..5597
/gene="biod"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase"
/protein_id="AAA23518.1"
/db_xref="GI:145428"
/translation="MSKRYEVTGDTVEYKTVASCALLQAAKRGYRTAGKRPVASS
KTEPEGLNSDALORNSLQIDYATNPPTFAEPSTPHLSQEGRPESLWMSAG
LRALEQOADMVTVGAGGWTPLSDTFEADWVQEOPLVLLVGVKLGGINHAMLTA
QVIOHAGLTLTAGWVANDVTPGKRHAEMTTLTMTIPRCWERSPGLOKIKKMPES
T"
BASE COUNT 1363 a 1554 c 1631 g 1245 t
ORIGIN 4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match 100.0%; Score 25; DB 2; Length 5793;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gacattgcgaagcacagaattta 25
|||||
Db 2032 GACATTGTCGCAAGCACAGAATTTA 2056

RESULT 6
LOCUS A38246 5872 bp DNA
DEFINITION Sequence 1 from Patent WO9408023.
ACCESSION A38246
VERSION A38246.1 GI:2294844
KEYWORDS
SOURCE
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL Patent: WO 9408023-A 1 14-Apr-1994;
LONZA AG (CH)
COMMENT
Other publication PL 308301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42095 951108
Other publication CZ 9500809 950913
Other publication FI 951547 950331
Other publication JP 8501694T 960227.
FEATURES
Source
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A-15/9"
1..96
/function="PROMOTER PTAC"
/evidence="experimental"
23..28
/function="PROMOTER PTAC"
45..49
/standard_name="PROMOTER PTAC"
/evidence="experimental"

```

```

RBS
105..119
/standard_name="BIOB RBS NO.9"
/evidence="experimental"
117..1157
/gene="BIOB"
117..1157
/gene="BIOB"
/codon_start=1
/transl_table=11
/number=1
/evidence="experimental"
/product="BIOTIN SYNTHASE"
/protein_id="CAA02325.1"
/db_xref="GI:2294845"
/translation="MAHPRRMTLSQVTELFKPLDILFEAQOVHROHFDPROVOST
LSIKTAGPEOCKYCPOSSRYKTGLFARLMEVQVLESARAKAAGSTRECMGAAW
KNHERDMYLEOMVGVKAMGLECMTGLTSESQORLANAGLDYNNHLDSPER
YGNITITRYQERLDLTKRVAGIKVCSGGIVIGLGVKRAAGLLQLANLPSPER
VPLNMLVKKGTPLADNDVDADFTRITAVRIMMPTSYVLSAGREOMNEQOAMC
FVAGANSIFYGCKLITTPNPEDEKDLQFLRKLGLNDPOQTAVLADGNEQOQRLGALMT
PDIDEYNNAAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/number=9
/product="PROTEIN"
/protein_id="CAA02326.1"
/db_xref="GI:2294846"
/translation="MATVVKOALIAAFGRAAHYFOHADLOROSADALLAMLPORRYT
HYLDAGCGPGMSRHRERHNOVATLSPMLVOAKRQDAADHTLADISLPLATA
TPEDLMSNLAVOMCGNSTALRELYRVPRGVAFITLVGSLPERHOAVADERP
HANRFLPDELEOSLNGVYOHIOPTLWFDLISAMRSIKIGATHLHGRDPRIL
TRSOQLRLQLAMPQOQGRYPLTYHLFLGVARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:
8-AMINO-7-OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/number=5
/evidence="experimental"
/product="DAPA SYNTHASE"
/protein_id="CAA02327.1"
/db_xref="GI:2294847"
/translation="MTTDLAFLDQRHIMHPTSMPLPYVYVSAGCELLISDGR
LYDGSSWMAIRGYNHPLQNAAKSQIDAMSHVPGTTHAPTELCRLVATTPOP
LBCYFLADSGSVAEVAAPKMALOYWQAKGEARQRTLFRRNGHGTGASVCDPDS
HSLIMKGYLPENLPAPOSRMDGEMBERDMVGFARLMAARHEIAVITEPIVCGAK
GMRWHPEMLKIRIKICDREGILLIADEIATGFGTGLFCAEHAETAPIILCGKAL
TGTGNTLSATLTREVAETISNGEGCGFMGPPMGNLPCAANASATLIESGDMO
QVADIEVOLREOLAPARDAEMVADVAVRIGAGVETHPYNNMALOKFVEQGWIRF
FGKLIYMPPTIILPQOLRLTAAVNRADDEFFCO"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/number=6

```

```

/evidence=experimental
/product="PROTEIN"
/protein_id="CAA02328.1"
/db_xref="GI:2294848"
/translation="MKLSNDLRDGLDPLRRHVEGNGWGDGDNISPHLAMPDVPAGTK
SEVVCYDPDPAFTGSGMMHVVYVNPATRTPOGFGGIVAMPDGYCOTRTDCKTG
YDGAAPKGETHRYLFTVHALDRIYDEGASGMVGFNVHFHSLASASTAAMS"
stem_loop
5583..5605
5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 gacattgcgaagtcacagaatta 25
Db      137 GACATTGCGCAAGTCACAGATT 161

RESULT 7
A38251      5872 bp      DNA      PAT      05-MAR-1997
LOCUS      Sequence 6 from Patent WO9408023.
ACCESSION  A38251
VERSION     A38251.1 GI:2294849
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM    Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE      BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL     Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694* 960227.
            Location/Qualifiers
FEATURES
Source      1..5872
            /organism="Escherichia coli"
            /strain="DSM498"
            /db_xref="taxon:562"
            /clone="PBO30A15-9"
            1141..1156
            /standard_name="BIOF RBS"
            1154..2308
            /gene="BIOF"
            1154..2308
            /gene="BIOF"
            /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
            /EC_number="2.3.1.47"
            /codon_start=1
            /transl_table=11
            /number=2
            /evidence=experimental
            /product="KAPA SYNTHASE"
            /protein_id="CAA02329.1"
            /db_xref="GI:2294850"
            /translation="MSMOEKINALDARRADALRRYPVAGACRMVLAADROYLNF
            SSNDYLGISHHPQILIRAWOGCAEQFGISGSGVSGSYVHQALEELAWLSTSYSA
            LFIISGFANAQAVIAAMAKEDRIAADRLSHASLLEASLSPSQILRRRAHNDVTHLAR
            LLASPCGQMVVTEGVEFSMDGDSAPLAEIOQVTOOHNGMLMVDADAHGTGIVGSGRG

```

```

SCMLQKVPPELLVYTFGKFGVSGAVALCSSTVADYLLQFARLLYSTSPQAQAL
RASLAVIRSDGDBAREKLEALITRFPAGVODLPFTLADSCSAIOPLIYDGNRSALOL
AEKLRQOGCWTAIRPPTVPAGTARLRLITFAHNEQDIDRLLEVLHGNG"
RBS      3030..3045
            /standard_name="BIOB RBS"
            3043..3753
            /gene="BIOB"
            3043..3753
            /gene="BIOB"
            /standard_name="DETHIOBIOTIN SYNTHASE"
            /EC_number="6.3.3.3"
            /codon_start=1
            /transl_table=11
            /number=4
            /evidence=experimental
            /product="PTB SYNTHASE"
            /protein_id="CAA02330.1"
            /db_xref="GI:2294851"
            /translation="MSKRYEVTGDTVEGKTVASCALLQAKAGYRTAGYKPVASG
            EKTPGGLNSDALAIORNSSLQDLYATVNPYTFEAPSPHISAOGRPIESLVSAG
            LRALEQADWVIVGAGGWFPLSDYTFEADWVTOEOLPVILVGVKLGICINAMLTA
            OVIOHAGITLIGWVANDYTPPKRKHAEVMTLTITMIPAPLIGELIPMLAENPENATGK
            YINLAEVDASTLGEFTSRL"
BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 gacattgcgaagtcacagaatta 25
Db      137 GACATTGCGCAAGTCACAGATT 161

RESULT 8
A93674      5872 bp      DNA      PAT      22-JAN-2000
LOCUS      Sequence 1 from Patent EP0798384.
ACCESSION  A93674
VERSION     A93674.1 GI:6741862
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM    Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch,O. and Brass,J.
TITLE      Biotechnological method of producing biotin
JOURNAL     Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
FEATURES
Source      1..5872
            /organism="Escherichia coli"
            /strain="DSM498"
            /db_xref="taxon:562"
            /clone="PBO30A-15/9"
            1..96
            /function="PROMOTER PTAC"
            /evidence=experimental
            45..49
            /standard_name="PROMOTER PTAC"
            105..119
            /evidence=experimental
            /standard_name="BIOB RBS NO.9"
            117..1157
            /gene="BIOB"
            117..1157
            /gene="BIOB"
            /codon_start=1

```

```

/transl_table=11
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHRPRMTLSQVTELEFEKPLLDLLFEAOVYRHOHFDPQOVST
LSTITGACPEDECKYCPQSSSRKGTGLEMERLEVEOVLESARKKAAGSTPGAAW
KNPHERMPYLEOMVGVKAMGLEACMTGLTSSQARLANAGIDYRHNHLDTPSPF
YGNITTTTVOERLDTLEKVDAGIKVCSGGIVGIGETVRKAGILLOLANLTPPSP
VPIMLVVKGTPLADNDVDADPFIIRIIVARIMPTSYRILSNGRONEGTQAMC
FMAGNSIFGCKLITTPNPEKDLQIFRKLGNLPOOTAVLAGDNEQOORLEQALMT
PDTEYVMAAL"
2284..2297
/standard_name="BIOC_RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVNNQALIAAGRAAHAECHADLQROSADALLMLPORYT
HYLDAGCGPQMSHMRERHAQVYALDLSPPMLVOARQKADHYLAGDIESLPATA
TFDLAMSNLAVQWGNLSTALRELYRVRKGVAAFTVLQSGSLPELHOMQVDRP
HANRELPDEPQSLNGVYOHHTOPTLTFMDALSAMSLKIGATHLHEGDRPIL
TRSOIQLQIQLAMPQOGRPLVYHLFLGVARE"
3742..3752
/standard_name="BIOA_RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONANONATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLADQRIHWPYISMTPLVPVPSAEGCELLISDGR
LVDMSSMAAHIGHYNHQIQAAMKSQIDAMSHVMGGITTHAPALECKIYAMTPOP
LECVPLADSGSAVEAAMKALQWQAKGEARORELTFRNGYHGTDFGAMSVCDPDS
MSHLMGYLPENLEFAPAPQSRMDGEMDERDVPARIMAAHHHEITAAVITPEYAG
GMRYHPMLTKRLKICDREGILLADFLATGCRGKI.FACEHAELADILICRAL
TGGMTLSATLTTRREVAETISNGEACGPHGTFKGNPLCAANAASLAIIESGMQO
OYADIEVOLREOLAPARADEMYADVAVIGATIGVETTHPVNMAALQKFEQGVIRP
FKGLIAMPYITILLPOQLORTAAVRAVODETFEQO"
5088..5100
/standard_name="ORFI_RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDQDKLPHRHVNGMGVGDGNTISPHLMDVPAQIK
SEVVCYDPDAPDGTSGMHWVYVNLPAITRVLPQSGSLVAMPDGVLOPTDCKGK
YDGAAPKGETHRTYFTVHALDIERIDVDEGASAMGFVNHPIHSLASASTTAMFS"
5583..5605
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
5583..5644
BASE COUNT 1318 a . 1552 c . 1695 g 1307 t

```

```

ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagtcacagaatta 25
Db 137 GACATTGTCGCAAGTCACAGAATTA 161

RESULT 9
A93679
LOCUS A93679 5872 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 6 from Patent EP0798384.
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 5872)
Birch, O. and Brass, J.
BIOLOGICAL METHOD OF PRODUCING BIOTIN
Patent: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)
FEATURES
Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A15-9"
1141..1156
/standard_name="BIOF_RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONANONATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSWQEKIINAALDARAADALRRRYPAVAGGRVIVADDOYLNF
SSNDYLGISHHPOIIRAMOGGAREGFGSGSGSHVSGT SYVHOLEBELEAEM/LGTISA
LLFTSGRANNAVITAAMAKEDRIADRLSHASLLEASLSPQLRFAHNDVTHLAR
LLASPCGQOQVYTBGVFSMDGSAPLAEIQVYTOQHNGLMWDADAGVIGEOGRG
SCWLQKVRPELVVTFEGKGGVSGAALVCSYADVLDLAFARHLITSTSPPAQOAL
RASLAVIRSDGDAEREKLAALITREFAGVODLPFLADSCSAIOPILVIGDNRALQ
AEKLRQOQCVWTAIRPPTVPAGTARLRILTAAHEMODIDRLLEVLAHNG"
3030..3045
/standard_name="BIOB_RBS"
3043..3753
/gene="BIOB"
3043..3753
/gene="BIOB"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYFVYGTQTEVNGKYVASCALLQANAAAGRTAGYFPVAGS
EKPEGRSNDALALQNRSSLDIATVNPVTFAEPTSPHIIISAEGRFESLVNSAG
LRALDEQADWLVLEAGGWFPLISDTFPAWVTOEOPLVILVGVKLGCLGNNAHMTA
QVIAQAGTLAGVANDVTPGKRRAEYVTLTRMIPAVILGEIPMLENPENATGK

```

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN YINIAPEVDASTGFTSRL"

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagtcacagaatta 25
 DB 137 GACATTGTGCAAGTCACAGAATTA 161

RESULT 10
 ARI01809 5872 bp DNA PAT 14-FEB-2001
 LOCUS
 DEFINITION Sequence 1 from patent US 6083712.
 ACCESSION ARI01809
 VERSION ARI01809.1 GI:12812607
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
 FEATURES Location/Qualifiers
 source 1..5872

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN /organism="Unknown"

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagtcacagaatta 25
 DB 137 GACATTGTGCAAGTCACAGAATTA 161

RESULT 11
 ARI01810 5872 bp DNA PAT 14-FEB-2001
 LOCUS
 DEFINITION Sequence 6 from patent US 6083712.
 ACCESSION ARI01810
 VERSION ARI01810.1 GI:12812608
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
 FEATURES Location/Qualifiers
 source 1..5872

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN /organism="Unknown"

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgtcgcaagtcacagaatta 25
 DB 137 GACATTGTGCAAGTCACAGAATTA 161

RESULT 12
 AE000180 11022 bp DNA BCT 01-DEC-2000
 LOCUS
 DEFINITION Escherichia coli K12 MG1655 section 70 of 400 of the complete
 genome.
 ACCESSION AE000180 U00096
 VERSION AE000180.1 GI:1786988
 KEYWORDS
 SOURCE Escherichia coli K12.
 ORGANISM Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE 1 (bases 1 to 11022)
 AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
 Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
 Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
 Mau, B. and Shaw, Y.
 TITLE The complete genome sequence of Escherichia coli K-12
 JOURNAL Science 277 (5331), 1453-1474 (1997)
 MEDLINE 97426617
 PUBMED 9278503

REFERENCE 2 (bases 1 to 11022)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecolli@genetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459

REFERENCE 3 (bases 1 to 11022)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecolli@genetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459

REFERENCE 4 (bases 1 to 11022)
 AUTHORS Plunkett, G. III.
 TITLE Direct Submission
 JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 This sequence was determined by the E. coli Genome Project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCHGR). The entire sequence was independently
 determined from E. coli K12 strain MG1655. Predicted open reading
 frames were determined using Genemark software, kindly supplied by
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, Ga.
 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
 have been correlated with genetic loci are being annotated with CG
 Site Nos., unique ID nos. for the genes in the E. coli Genetic
 Stock Center (CGSC) database at Yale University, kindly supplied by
 Mary Berlyn. A public version of the database is accessible
 (<http://cgsc.biology.yale.edu>). Annotation of the genome is an
 ongoing task whose goal is to make the genome sequence more useful
 by correlating it with other data. Comments to the authors are
 appreciated. Updated information will be available at the E. coli
 Genome Project's World Wide Web site
 (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and
 its annotations are periodically updated; this is version M54. No
 sequence changes. Annotation updates: updated gene identifications
 and products; all new functional assignments courtesy of Monica
 Riley; added promoters, protein binding sites, and repeated
 sequences described in reference 1. The unique numeric identifiers
 beginning with a lowercase 'b' assigned to each gene (protein- or
 RNA-encoding) are now designated as gene synonyms instead of
 labels. This should allow them to be searched for in Entrez as gene
 names.

FEATURES
 source Location/Qualifiers
 1..11022 /organism="Escherichia coli K12"
 /strain="K12"
 /sub_strain="MG1655"


```

VERSION      AE005258.1  GI:12513751
KEYWORDS
SOURCE       Escherichia coli O157:H7 EDL933.
ORGANISM     Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
              Escherichia.
REFERENCE    1 (bases 1 to 13501)
AUTHORS      Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
              Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
              Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
              Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
              Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
              Welch,R.A. and Blattner,F.R.
TITLE        Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
JOURNAL      Nature 409 (6819), 529-533 (2001)
MEDLINE      21074935
PUBMED       11206551
REFERENCE    2 (bases 1 to 13501)
AUTHORS      Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
              Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
              Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
              Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
              Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
              Welch,R.A. and Blattner,F.R.
TITLE        Direct Submission
JOURNAL
JOURNAL      Submitted (22-OCT-2000) Laboratory of Genetics, University of
              Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
source       1. 13501
              /organism="Escherichia coli O157:H7 EDL933"
              /strain="EDL933"
              /serotype="O157:H7"
              /db_xref="taxon:155864"
              /note="enterohemorrhagic"
              <1. 7576
              /note="O-island #36; Region of the EDL933 chromosome not
              homologous to E. coli K-12 MG1655; Cryptic prophage
              CP-933K; includes one copy of the 13 bp direct repeat that
              flanks the prophage"
              66. 665
              /gene="lomk"
              /note="20981"
              /gene="lomk"
              /note="20981"
              /function="putative membrane; Other or unknown (Phage or
              Prophage Related)"
              /note="Residues 1 to 199 of 199 are 72.86 pct identical to
              residues 1 to 199 of 199 from Genpept 118 :
              g117532789|gb|AAE63231.1|AF151091_2 (AF151091) Lom
              [prophage P-E1bA]"
              /transl_start=1
              /transl_table=1
              /product="putative outer membrane protein of prophage
              CP-933K"
              /protein_id="AAG55137.1"
              /db_xref="GI:12513752"
              /translation="MRVCAIITSAICLAIVGVANASEHOSTLSAGYLAHASTDAG
              SDDLNGINVKRYEFTDLGLITFSYANAEDEKTHYSDTRHEDYVRRKRWISVAG
              PSRVNEMFSAYAGAVASRVSPSGDIFRVNTKRTKTHDVLITGSDARYSMTSLAM
              GAGYQNPTESEAVDAVAVEXSGSDMRTDGFIVGVYK"
              719. 2041
              /gene="20982"
              /gene="20982"
              /function="putative structure; Structural component (Phage
              or Prophage Related)"
              /note="Residues 164 to 440 of 440 are 68.79 pct identical
              to residues 381 to 645 of 645 from Genpept 118 :
              g114585436|gb|AAD25464.1|AF125520.59 (AF125520) putative
              tail fiber protein [Bacteriophage 933W]"
              /codon_start=1
              /transl_table=1
              /product="putative tail component of prophage CP-933K"

gene
CDS
              /protein_id="AAG55138.1"
              /db_xref="GI:12513753"
              /translation="MGMAAVQISGVKLGKAGKQKPIONCTIQKARNSTXYVYVTVASE
              NPDEAGRYSMDEVGYOYSVTLLEGFPPSHAGTISVREDSPGLNDFLGMTEDDK
              PEALRRFEQWVEEARHAHEAKKNAGEETSAKNAGISASKAEKSNANADTSADABE
              SARQAASEASAKSEESASSASAEAAKASLSQASADATLESKTAESAGNARA
              TTSTKRAESASQASBOSRTIAEDAVNRITPTVYVGGPGPGPGAPGPGKDGKE
              RGDTPAGATGEGRGCGDTPAGPGGPGGDEGREGELTGAGPGGKGGTGAAGA
              GGKGGKGTGAAGAPGATGPGPGKPKDPPETQIRFLGEMRIETNSYSPFGTGCALI
              TGLTLPDPAQVQGMFOHLDVRRGDEGPDWDVKKLDEVGSDTGTFE"
              2004. 2312
              /gene="20984"
              2004. 2312
              /gene="20984"
              /function="orf; Other or unknown (Phage or Prophage
              Related)"
              /note="Residues 1 to 102 of 102 are 98.03 pct identical to
              residues 28 to 129 of 129 from Genpept 118 :
              g114585437|gb|AAD25465.1|AF125520.60 (AF125520)
              hypothetical protein [Bacteriophage 933W]"
              /codon_start=1
              /transl_table=11
              /product="unknown protein encoded by prophage CP-933K"
              /protein_id="AAG55139.1"
              /db_xref="GI:12513754"
              /translation="MKMAVIOAELENDNNILPKLMOSLGGCGKHDCENGSLTAOLR
              LGPADLSEDENGIIPEDRITTVYIILDADKKQIQCVYRPLQIIRADGTENIGMK"
              2489. 3469
              /gene="20985"
              2489. 3469
              /gene="20985"
              /function="orf; Other or unknown (Phage or Prophage
              Related)"
              /note="Residues 13 to 326 of 326 are 55.73 pct identical
              to residues 18 to 331 of 336 from Genpept 118 :
              g116960367|gb|AAE63527.11 (AF170176) hypothetical protein
              predicted by Glimmer [Salmonella typhimurium LT2]"
              /codon_start=1
              /transl_table=11
              /product="unknown protein encoded by prophage CP-933K"
              /protein_id="AAG55140.1"
              /db_xref="GI:12513755"
              /translation="MSPIRITMFPNSVNTVOSSPCQVTSFAGEKELAVIDEKPPILF
              QFEPPEPKKKDEVPIYTKQHPPLIDNTNNAKRFESPMIGIIPVDCGFSVQKAFS
              KLERDENWIIYREDVDSMTDRKLSIDYIHDIICEQRLRTDKRDETLNLEKEIK
              KESKADQSLISMAKKRHWMPDFRNALILAGELFCTYWTKNHGISFGGCIYLD
              MDMLILGKGTIYAPDGISMHYDRDRDSVINENSAIIVRSNHPALLLEGLSPMSKVD
              AHRYVGLGKGVKKYKPNFPLNHNHFCDFIEFNHPIIMNTSQTCSM"
              3530. 4522
              /gene="20986"
              3530. 4522
              /gene="20986"
              /function="orf; Other or unknown (Phage or Prophage
              Related)"
              /note="No significant matches"
              /codon_start=1
              /transl_table=11
              /product="unknown protein encoded by prophage CP-933K"
              /protein_id="AAG55141.1"
              /db_xref="GI:12513756"
              /translation="MKPIQSQNFNSAPAGYSAPTPAPNAENAVADYLDIGKRPL
              SAADLSNVEESVTRAVHSRLIDQHYDMIGNTVLDALSSQTFRAVSGTINNEK
              VHIGCIKRYNEVELNEESSVKIDDIQSLILNLEEDVGQEPFIPICENAGENDNEPY
              VSFVSVPDSDSEMPQWQGLIHEIITHVTGSSDSCDSNIEILGPTETIARVADELG
              MSVPDPKGYAEPEERHRLRLNMLALQQAAMHEENRAFFERLGTISIRYBASDFPT
              EYSAVSNIGYGIQCHDPGLAINDMLQDANQIQLYHGAPYIFRTGVDVKKHMQQ"
              5350. 6231
              /gene="20989"
              5350. 6231
              /gene="20989"
              /function="orf; Other or unknown (Phage or Prophage
              Related)"

```


PUBMED 1113432
2 (bases 1 to 5526)
AUTHORS Entcheva, P., Liebl, W. and Streitt, W.R.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES
source
1. 5526
/organism="uncultured bacterium pcosHE2"
/db_xref="taxon:143797"
/clone="pcosHE2"
/note="unknown organism, cosmid clone derived from
environmental consortium"
complement(52..528)
/note="ORF1"
/codon_start=1
/transl_table=1
/product="hypothetical 17.1 kDa protein in modC-bioA
intergenic region"
/protein_id="AA60577.1"
/db_xref="GI:12620125"
/translation="MKLISNDLRDGDKLPHRHVFMGMDGDNISPHLAMDVPAGTK
SEVYTCYDADPAGSGMHMVVNLPAADTRVLPQFGSGLVAMPDGLVOTRTDFGKAG
YDGAAPKGETHRVITVHALDIERLDVDEGASGAMVGFVHFHSLASITAMS"
complement(587..1876)
/gene="bioA"
complement(587..1876)
/gene="bioA"
/note="7,8-diaminopelargonic acid
synthetase-aminotransferase"
/codon_start=1
/transl_table=1
/product="DAPA-aminotransferase bioA"
/db_xref="GI:12620126"
/translation="MTTDLAPDQRHIMHPTSMTPSPVYVVSAGCELLISDGR
LVDSMSWMAIHGYNHNPQLNAAMKSDIDAMSHVFGITTHAPATELRKLIYAMTPO
LACVGLADSGVAEVAAMKALQYQAKGEARQFLTPRNGYHGTDFGMSVCDDNS
MSLWKGFLPENLPAPOSMDGEMDEDMVGFARMAAHHEITAAVITIPYIYAG
GRRHYHEWLKRIKICDREGILLADELAFRGCTGKLPACNHAETADTLCLAKL
TGGTMTLSATITREAVLTISNGEACGCMHGTPTGNCPLACAAASLALIESGMOO
QYADIEVOLREOLAPARDAENADVRLGATIGVETTPVNNMALOKFFVBOGWIRP
FEKLIYMPPIYILPQOLRIITAAVNAVODETFPCQ"
1963..3003
/gene="bioB"
1963..3003
/gene="bioB"
/codon_start=1
/transl_table=1
/product="biotin synthase bioB"
/protein_id="AAG60579.1"
/db_xref="GI:12620127"
/translation="MAHRPMTLSQVTFLEPKILLILLFEAOYVHRHPPROYST
LLSTGACPECKCYCPSSRYKGTLEARELMEYQVYESRKAKASTFRGGAAM
KNPNRMDPYLEOMYQGYKDGLEACHTLGLTSSQAKRLAGIDYNNHLDTPSP
YGNITTTTTOERLDTLEKVDAGIKVCSGGIVGIGETVKDRAILLQLANLPTPPS
VPIINLVKGTPLADNDVDAPDFETRTI AAVRIIMPVSYLISAGSEOMNEQOAMC
FMAGANSIFGYCKLITPNPEEDKDLQIFRKILGNPQOTAVLAGDNEQQQLBQALMT
PDTDEYNAAL"
3000..4154
/gene="bioF"
3000..4154
/gene="bioF"
/note="8-amino-7-oxononanoate synthase"
/codon_start=1
/transl_table=1
/product="KAPA synthetase bioF"
/protein_id="AAG60580.1"
/db_xref="GI:12620128"
/translation="MSWQEKINAALARRAADALRRRYPYAAGAGFMLVADBOYLNF
SSNDYLGSHHPQIIRAMQOAGQEGIGSGVSGVSVVHQALEBELAEMWIGYSA
LLFTSGAANAQAVIIRAMAKEDRIADRLSHASLLEAASISPSQLRFPANDYTHLAR
LLASPGQOGLVIT EGVFSMDGSAPIALETIOQVITQCHNGMLMDADHAGTGTGEGRC

SCNLOKVPKELLVYTFGKRGVSGAIVCSVADYLLQFARHLITSTMPQAQAL
RASLAVITRDSGDARREKIVLSLHFRAGVDDUPTLADSCSLQPLIVDINSRALQ
AEKLRQOCVWTAIRPPVPAFTARLRTITAAHEMODIDLLEVLHNG"
4141..4896
/gene="bioC"
4141..4896
/note="reaction step prior to synthesis of pimeloyl-CoA"
/codon_start=1
/transl_table=1
/product="biotin biosynthesis protein bioC"
/protein_id="AAG60581.1"
/db_xref="GI:12620129"
/translation="MATVNCQAIARFAAGRAALYEOHADLOROSADYLLAMPKRYT
HYLDAGGPGMTWRHRRHRAQVYALDLSPPMYQAROKMAHYLAGDIESIPLATA
TEPLAWSNLAQMGKGNSTALRELYRVRSKYVAFITLVGSLPELHQAWQAVDERP
HANRPLPEDELSLNGVTHQHIQPTTLWFDLMSRLKIGATHLEGRDPRIL
TRSLQRLQALAMPQOGRYPVLYHLFGVIARE"
4883..5526
/gene="bioD"
4883..5526
/gene="bioD"
/note="DTB synthetase"
/codon_start=1
/transl_table=1
/product="delhiobiotin synthetase bioD"
/protein_id="AAG60582.1"
/db_xref="GI:12620130"
/translation="MIVSKRYFPGDTEFGKTVASCALLQAAKAGYRAGKPPAS
GSEKTPGINSQALAQKNSIQDQATVATNPPTPEPTSPHISAOGRPISSVMS
SGIRALEQADWLVESAGGWTPLSDTFYFAWVQEDPLVTVGVKICINIAML
TQAQIQAGLILGAWANDVTPPKRIAEVITVITRIPADLIGELPMAENBE"
1507 a 1507 c 1567 g 1178 t
BASE COUNT 1274 a 1507 c 1567 g 1178 t
ORIGIN
Query Match 84.0% Score 21; DB 2; Length 5526;
Best Local Similarity 100.0%; Pred. No. 0.0045;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 ttgtgcagtcacagaatta 25
|||||
Db 1987 TTGTGCAAGTCACAGATT 2007
RESULT 16
AF248314
LOCUS 8227 bp DNA BCT 24-JAN-2001
DEFINITION uncultured bacterium pcosAS1 urocanase-like protein (hutu) gene,
partial cds; histidine ammonia-lyase-like protein (hutu), DAPA
aminotransferase bioA (bioA), biotin synthase bioB (bioB), 7-KAPA
synthetase (bioF), biotin biosynthesis bioC-like protein (bioC),
and delhiobiotin synthase bioD (bioD) genes, complete cds; ABC
transporter-like protein (elsA) gene, partial cds; and unknown
gene.
ACCESSION AF248314
AF248314.1 GI:12407610
VERSION 1 (bases 1 to 8227)
KEYWORDS uncultured bacterium pcosAS1.
SOURCE Bacteria; environmental samples.
ORGANISM uncultured bacterium pcosAS1.
REFERENCE 1 (bases 1 to 8227)
AUTHORS Entcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streitt, W.R.
TITLE Direct Cloning from Enrichment Cultures, a Reliable Strategy for
Isolation of Complete Operons and Genes from Microbial Consortia
JOURNAL Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
PUBMED 1113432
REFERENCE 2 (bases 1 to 8227)
AUTHORS Entcheva, P., Liebl, W. and Streitt, W.R.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2000) Mikrobiologie und Genetik, Universitaet
Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES
Location/Qualifiers

```

source
1. .8227
/organism="uncultured bacterium pCosAS1"
/db_xref="taxon:143793"
/clone="pCosAS1"
/note="unknown organism, cosmid clone derived from
environmental consortium"
<1. .502
/gene="hutu"
<1. .502
/gene="hutu"
/note="putative urocanate hydratase"
/codon_start=2
/transl_table=11
/product="urocanase-like protein"
/protein_id="AAG53593.1"
/db_xref="GI:12407618"
/translation="PRRISFQGLPARICWVGLBWRQKLGIAFNEMVRGSEVAPIV
GRDHDGSSVASPNRETERAMQSDVSDWPLNALNTASGATWVSLHGGVGKGF
SOHAGMVICDGTDDAARLARLVHNDPATGVMHADAGYDIALDCAKEQSLNPMIA
ATQGGH"
502. .2031
/gene="huth"
502. .2031
/gene="huth"
/note="putative"
/codon_start=1
/transl_table=11
/product="histidine ammonia-lyase-like protein"
/protein_id="AAG53586.1"
/db_xref="GI:12407611"
/translation="MNALITPTGTLTALQROVMOOPLQLTDSAHBAINDVACVE
ALVAGRTAYGINTGFGLLAOTRITTHLENQKSLVSHAAGVEPLDDIDVLMVY
LKINSLANGFSGIRLSVIOALIALVNAVSVDPKSGVSGDLAPLHHSLSLTLGE
GKARYGEMLPATALOKAGLAPVTLAKREGIALINGQASTAFELRGLFEADHFLAS
AVVCGALITTEAVLGSRRPDDARIEVRQGRQIDAAALFRVLTDTSAIAHSNHCCK
VODPYSLRGPOVWGACLTOMROVAEVLVESNADNPDLVFAENEVAFGNFAEP
VAMADNLALIAETAGLISERRIALMCKHNSQLPPLVNGVNSGFMIAOVTAAL
ASENKGICHPHTTSYDKPCRSANEDIVSNAPARGRLMEAGTSGVLAVENIAC
OGADLRDGLTSSPLLEQARQSCGEQVAHDDRFPAPIEATISLNKGSVLGLLPAP
L"
complement(2130. .2606)
/note="ORF1, similar to hypothetical 17.1 kDa protein in
modc-b10a intergenic region"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAG53587.1"
/db_xref="GI:12407612"
/translation="MKLISQDLRDGDKPLRHVNGMGYEGDNI SPHLAMDEPVGTK
SFVVCYDDPAPDPSGMMHVIYANI PADTRVLPQSGSSIVALPGCAVETTRDPCKAG
YCGAAPPKGETHRYITVHALVERIEVDDEASGAMVGNVHFLGSASITAMS"
complement(2651. .3940)
/gene="bioA"
complement(2651. .3940)
/note="7.8-diamino-pelargonic acid aminotransferase"
/codon_start=1
/transl_table=11
/product="DAPA aminotransferase BioA"
/db_xref="GI:12407613"
/translation="MTTDDLAFDRKHQIIMHPTSMOPLPVYPUVEAHGELTLASGEQ
LVNGSSMMAATIHGYNHPRPLNAMKAOIDRMHVMGCGITTHQPAVDLCRGLVAMPET
LECVFPADSGSAVEYAKMAIOYHNHAKESQRPPTTPNGYHGFQFGMSYCDGAS
MHSVMKGYLPENLFAPQSRFDGANDMDWGLAPLMAHHEIATAVILEYVOCAG
GMSMYHPWMLKRIIRMCQDRGILLIADETATGFGTKLFAEHAGITPDIICLGAL
TGGTMTSATLTTRQVAETISNGEACGFMHGPFGNPLACAVASESLALSSGEMOD
OYVAIESOLOEELAVARDESEFADVAVLAIQVETTHPVNMAALQRFVAVGVWVR
FGKLTLMPPYVIRPDQSLKLTAVVADVAVQAPAHFTR"
4029. .5069
/gene="b10B"
4029. .5069
/gene="b10B"
/ gene=
CDS
/ gene="b10B"

gene
1512 a 2642 c 2605 g 1453 t 15 others
BASE COUNT

```

ORIGIN

Query Match 72.0% Score 18; DB 2; Length 8227;
 Best Local Similarity 100.0%; Pred. No. 0.36;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattcgcgaagtcac 18
 ||||||||||||||||
 DB 4049 GACATTGCGCAAGTCAC 4066

RESULT 17

AC087305/c
 LOCUS AC087305
 DEFINITION Homo sapiens chromosome 18 clone RP11-238P13 map 18, LOW-PASS
 SEQUENCE SAMPLING.

AC087305 32767 bp DNA HTG 23-DEC-2000
 AC087305.1 GI:11990696
 HTG: HTGS_PHASED.
 ORGANISM human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 32767)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., and Lander, E.
 TITLE Homo sapiens chromosome 18, clone RP11-238P13
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 32767)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
 Baran, N., Bastien, V., Boguslavsky, L., Bouckhaert, B., Brown, A.,
 Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J., S.,
 Dodge, S., Fano, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,
 Gardyna, S., Glade, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hages, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Karatas, A., Labocque, K., Lamazares, R., Landers, T.,
 Lehoczy, J., Levine, R., Liu, G., McClean, C., MacDonald, P.,
 Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
 McPheters, R., Meldrum, D., Meneus, L., Minova, T., Mlenga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H.,
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Phukhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
 Plumbach, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roselli, M.,
 Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P.,
 Sougnez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
 Travers, M., Travis, N., Triglilio, J., Vassiliev, H., Viel, R., Vo, A.,
 Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J.,
 Zembek, L., Zimmer, A. and Zody, M.

TITLE
 JOURNAL
 COMMENT
 Submitted (23-DEC-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 ALL repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: MIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: LI2071
 Center clone name: 238_P_13

* NOTE: This record contains 40 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone

* will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.
 1
 708 807: gap of 100 bp in length
 808 1527: contig of 720 bp in length
 1528 1627: gap of 100 bp
 1628 2348: contig of 721 bp in length
 2349 2448: gap of 100 bp
 2449 3157: contig of 709 bp in length
 3158 3257: gap of 100 bp
 3258 3973: contig of 716 bp in length
 3974 4073: gap of 100 bp
 4074 4799: contig of 726 bp in length
 4800 4899: gap of 100 bp
 4900 5617: contig of 718 bp in length
 5618 5717: gap of 100 bp
 5718 6450: contig of 733 bp in length
 6451 6550: gap of 100 bp
 6551 7259: contig of 709 bp in length
 7260 7359: gap of 100 bp
 7360 8066: contig of 707 bp in length
 8067 8166: gap of 100 bp
 8167 8907: contig of 741 bp in length
 8908 9007: gap of 100 bp
 9008 9716: contig of 709 bp in length
 9717 9816: gap of 100 bp
 9817 10532: contig of 716 bp in length
 10533 10632: gap of 100 bp
 10633 11354: contig of 722 bp in length
 11355 11454: gap of 100 bp
 11455 12166: contig of 712 bp in length
 12167 12266: gap of 100 bp
 12267 12976: contig of 710 bp in length
 12977 13076: gap of 100 bp
 13077 13789: contig of 713 bp in length
 13790 13889: gap of 100 bp
 13890 14617: contig of 728 bp in length
 14618 14717: gap of 100 bp
 14718 15452: contig of 735 bp in length
 15453 15552: gap of 100 bp
 15553 16276: contig of 724 bp in length
 16277 16376: gap of 100 bp
 16377 17111: contig of 735 bp in length
 17112 17211: gap of 100 bp
 17212 17925: contig of 714 bp in length
 17926 18025: gap of 100 bp
 18026 18746: contig of 721 bp in length
 18747 18846: gap of 100 bp
 18847 19563: contig of 723 bp in length
 19570 19669: gap of 100 bp
 19670 20384: contig of 715 bp in length
 20385 20484: gap of 100 bp
 20485 21190: contig of 706 bp in length
 21191 21290: gap of 100 bp
 21291 22028: contig of 738 bp in length
 22029 22128: gap of 100 bp
 22129 22841: contig of 713 bp in length
 22842 22941: gap of 100 bp
 22942 23653: contig of 712 bp in length
 23654 23753: gap of 100 bp
 23754 24478: contig of 725 bp in length
 24479 24578: gap of 100 bp
 24579 25305: contig of 727 bp in length
 25306 25405: gap of 100 bp
 25406 26143: contig of 738 bp in length
 26144 26243: gap of 100 bp
 26244 26968: contig of 725 bp in length
 26969 27068: gap of 100 bp
 27069 27785: contig of 717 bp in length
 27786 27885: gap of 100 bp
 27886 28584: contig of 699 bp in length
 28585 28684: gap of 100 bp

```

* 28685 29467: contig of 783 bp in length
* 29468 29567: gap of 100 bp
* 29568 30292: contig of 725 bp in length
* 30293 30392: gap of 100 bp
* 30393 31117: contig of 725 bp in length
* 31118 31217: gap of 100 bp
* 31218 31944: contig of 727 bp in length
* 31945 32044: gap of 100 bp
* 32045 32767: contig of 723 bp in length.
Location/Qualifiers
1..32767
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-238P13"
/clone_lib="RPC1-11 Human Male BAC"

```

```

BASE COUNT      8045 a 5670 c 5396 g 9330 t 4326 others
ORIGIN

```

```

Query Match      64.0%; Score 16; DB 77; Length 32767;
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      10 gcaagtcacagaatta 25
        |||
DB 30730 GCAAGTCACAGAATTA 30715

```

```

RESULT 18
LOCUS      AC021110 155172 bp DNA HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-238P13, WORKING DRAFT
ACCESSION  AC021110
VERSION     AC021110.3 GI:8439978
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 155172)
            Waterston,R.H.
            The sequence of Homo sapiens clone
            Unpublished
            2 (bases 1 to 155172)
            Waterston,R.H.
            Direct Submission
            Submitted (14-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT    On Jun 10, 2000 this sequence version replaced gi:7021826.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information
Center project name: H_NH0238P13
Summary Statistics
Sequencing vector: p13; 73%
Chemistry: Dye-primer ET; 73% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 15162 bases at least Q40
Consensus quality: 15244 bases at least Q30
Consensus quality: 153249 bases at least Q20
Insert size: 177000; agarose-ftp
Insert size: 154672; sum-of-ctigs
Quality coverage: 4.08 in Q20 bases; agarose-ftp
Quality coverage: 4.69 in Q20 bases; sum-of-ctigs

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
9265: contig of 9265 bp in length
9266 9365: gap of unknown length
9366 26504: contig of 17139 bp in length
26505 26605: gap of unknown length
44300: contig of 17696 bp in length
44301 44400: gap of unknown length
44401 62250: contig of 17850 bp in length
62251 62350: gap of unknown length
62351 105744: contig of 43394 bp in length
105745 105844: gap of unknown length
105845 155172: contig of 49328 bp in length.
Location/Qualifiers
1..155172
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/clone="RP11-238P13"

```

```

FEATURES
source
misc_feature 1..9265
            /note="assembly_name:Contig1"
            /db_xref="taxon:9606"
            /clone="RP11-238P13"
misc_feature 9366..26504
            /note="assembly_name:Contig2"
            /note="assembly_name:Contig3"
            /note="assembly_name:Contig4"
            /note="assembly_name:Contig5"
            /note="assembly_name:Contig6"
misc_feature 26605..44300
            /note="assembly_name:Contig6"
misc_feature 44401..62250
            /note="assembly_name:Contig6"
misc_feature 62351..105744
            /note="assembly_name:Contig6"
misc_feature 105845..155172
            /note="assembly_name:Contig6"
BASE COUNT 51543 a 28385 c 28043 g 46698 t 503 others
ORIGIN

```

```

Query Match      64.0%; Score 16; DB 66; Length 155172;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      10 gcaagtcacagaatta 25
        |||
DB 11929 GCAAGTCACAGAATTA 11944

```

```

RESULT 19
LOCUS      AC041018 156184 bp DNA HTG 22-MAY-2000
DEFINITION Homo sapiens chromosome 4 clone RP11-478020 map 4, WORKING DRAFT
ACCESSION  AC041018
VERSION     AC041018.2 GI:8016453
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 156184)
            Birren,B., Linton,L., Nussbaum,C. and Lander,E.
            Homo sapiens chromosome 4, clone RP11-478020
            Unpublished
            2 (bases 1 to 156184)
            Birren,B., Linton,L., Nussbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barina,N., Bastien,V., Bedalov,F.,
            Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,

```

TITLE
JOURNAL
COMMENT

Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collins, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Gadigan, J., Gardina, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Larroque, K., Lamas, R., Landers, T., Lehoczy, J.,
Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrum, J., Menus, L., Miho, T., Miranda, C., Mienna, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Olliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisanli, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 22, 2000 this sequence version replaced g1:7534220.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

Project Information

Center project name: 10577

Center clone name: 478_O_20

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 147958 bases at least Q40

Consensus quality: 151874 bases at least Q20

Consensus quality: 153413 bases at least Q20

Insert size: 15700; agarose-1p

Insert size: 154684; sum-of-ctigs

Quality coverage: 4.8 in Q20 bases; agarose-1p

Quality coverage: 4.8 in Q20 bases; sum-of-ctigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

* 1 1351: contig of 1351 bp in length
* 1352 1451: gap of 100 bp
* 1452 4369: contig of 2918 bp in length
* 4370 4469: gap of 100 bp
* 4470 8873: contig of 4404 bp in length
* 8874 8973: gap of 100 bp
* 8974 13167: contig of 4194 bp in length
* 13168 13267: gap of 100 bp
* 13268 18989: contig of 5722 bp in length
* 18990 19089: gap of 100 bp
* 19090 24989: contig of 5900 bp in length
* 24990 25089: gap of 100 bp
* 25090 33460: contig of 8371 bp in length
* 33461 33560: gap of 100 bp
* 33561 42328: contig of 8768 bp in length
* 42329 42428: gap of 100 bp
* 42429 50674: contig of 8246 bp in length
* 50675 50774: gap of 100 bp
* 50775 59862: contig of 9088 bp in length
* 59863 59962: gap of 100 bp

```

FEATURES

source

```

* 59963 71699: contig of 11737 bp in length
* 71700 71799: gap of 100 bp
* 71800 82583: contig of 10784 bp in length
* 82584 82683: gap of 100 bp
* 82684 95072: contig of 12389 bp in length
* 95073 95172: gap of 100 bp
* 95173 108949: contig of 13777 bp in length
* 108950 109049: gap of 100 bp
* 109050 126668: contig of 17619 bp in length
* 126669 126769: gap of 100 bp
* 126769 156184: contig of 29416 bp in length.
  Location/Qualifiers
    1..156184
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="4"
      /map="4"
      /clone="RP11-478020"
      /clone_11b="RPC1-11 Human Male BAC"
    1..1351
      /note="assembly-fragment"
    1452..4369
      /note="assembly-fragment"
    4470..8873
      /note="assembly-fragment"
    clone_end:SP6
      /note="assembly-fragment"
    vector_side:right"
    8974..13167
      /note="assembly-fragment"
    13268..18989
      /note="assembly-fragment"
    19090..24989
      /note="assembly-fragment"
    25090..33460
      /note="assembly-fragment"
    33561..42328
      /note="assembly-fragment"
    42429..50674
      /note="assembly-fragment"
    50775..59862
      /note="assembly-fragment"
    59963..71699
      /note="assembly-fragment"
    71800..82583
      /note="assembly-fragment"
    82684..95072
      /note="assembly-fragment"
    clone_end:T7
      /note="assembly-fragment"
    vector_side:right"
    95173..108949
      /note="assembly-fragment"
    109050..126668
      /note="assembly-fragment"
    126769..156184
      /note="assembly-fragment"

```

```

BASE COUNT  51507 a 26941 c 26653 g 49578 t 1505 others
ORIGIN

```

```

Query Match      64.0%; Score 16; DB 71; Length 156184;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 10 gcaagtcacagaatta 25
      |||
Db 123532 GCAAGTCACAGAATTA 123517

```

```

RESULT 20
AP001390/c      DNA      HTG      30-MAY-2000
LOCUS          Homo sapiens chromosome 18 clone RP11-635B11 map 18q23, WORKING
DEFINITION
DRAFT SEQUENCE, 21 unordered pieces.

```


ACCESSION AP001390 GI:8117301
 VERSION AP001390.2
 KEYWORDS HTGS_PHASE1; HTGS_DRAFT;
 SOURCE Homo sapiens DNA, clone:RP11-635B11.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 174815)
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 Homo sapiens 174,815 genomic DNA of 18q23
 Published Only in Database (2000) In press
 2 (bases 1 to 174815)
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 Direct Submission
 Submitted (14-MAR-2000) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
 Japan (E-mail:hattori@gscc.riken.go.jp,
 URL:http://hqp.gsc.riken.go.jp/, Tel:81-42-778-9923,
 Fax:81-42-778-9924)
 On May 30, 2000 this sequence version replaced gi:7259325.
 ----- Genome Center
 Center: RIKEN Genomic Sciences Center(GSC)
 Center code: RIKEN
 Web site:http://hqp.gsc.riken.go.jp/
 Contact: hattori@gscc.riken.go.jp
 ----- Project Information
 Center project name: Humdraft18
 Center clone name: RP11-635B11
 ----- Summary Statistics
 Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 159194 bases at least Q40
 Consensus quality: 167087 bases at least Q30
 Consensus quality: 170693 bases at least Q20
 Insert size: 172815; sum-of-contigs
 Quality coverage: 4.94x in Q20 bases; sum-of-contigs

 NOTE: This is a 'working draft' sequence. It currently consists of
 21 contigs. The true order of the pieces is not known and their
 order in this sequence record is arbitrary. Gaps between the
 contigs are represented as runs 'N', but the exact sizes of the gaps
 are unknown. This record will be updated with the finished sequence
 as soon as it is available and the accession number will be
 preserved

***** arbitrary. Gaps between the contigs are represented as
 ***** runs of N, but the exact sizes of the gaps are unknown.
 ***** This record will be updated with the finished sequence
 ***** as soon as it is available and the accession number will
 ***** be preserved.
 1 58823: contig of 58823 bp in length
 58824 58923: gap of 100 bp
 58924 73243: contig of 14320 bp in length
 73244 73343: gap of 100 bp
 73344 83572: contig of 10229 bp in length
 83573 83672: gap of 100 bp
 83673 94547: contig of 10875 bp in length
 94548 94647: gap of 100 bp
 94648 104597: contig of 9950 bp in length
 104598 104697: gap of 100 bp
 104698 114347: contig of 9650 bp in length
 114348 114447: gap of 100 bp
 114448 122080: contig of 7633 bp in length
 122081 122180: gap of 100 bp
 122181 128465: contig of 6285 bp in length
 128466 128565: gap of 100 bp
 128566 135791: contig of 7226 bp in length
 135792 135891: gap of 100 bp
 135892 142934: contig of 7043 bp in length
 142935 143034: gap of 100 bp
 143035 147999: contig of 4965 bp in length
 148000 148099: gap of 100 bp
 148100 151734: contig of 3635 bp in length
 151735 151834: gap of 100 bp
 151835 155576: contig of 3742 bp in length
 155577 155676: gap of 100 bp
 155677 159193: contig of 3517 bp in length
 159194 159293: gap of 100 bp
 159294 161359: contig of 2066 bp in length
 161360 161459: gap of 100 bp
 161460 163515: contig of 2056 bp in length
 163516 163615: gap of 100 bp
 163616 166614: contig of 2999 bp in length
 166615 166714: gap of 100 bp
 166715 169056: contig of 2342 bp in length
 169057 169156: gap of 100 bp
 169157 171527: contig of 2371 bp in length
 171528 171627: gap of 100 bp
 171628 173248: contig of 1621 bp in length
 173249 173348: gap of 100 bp
 173349 174815: contig of 1467 bp in length.
 Location/Qualifiers
 1. 174815
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18q23"
 /clone="RP11-635B11"
 1. 58823
 /note="assembly-fragment"
 58924. 73243
 /note="assembly-fragment"
 73344. 83572
 /note="assembly-fragment"
 83673. 94547
 /note="assembly-fragment"
 94648. 104597
 /note="assembly-fragment"
 104698. 114347
 /note="assembly-fragment"
 114448. 122080
 /note="assembly-fragment"
 122181. 128465
 /note="assembly-fragment"
 128566. 135791
 /note="assembly-fragment"
 135892. 142934
 /note="assembly-fragment"

Sequence updated (26-May-2000).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 21 contigs. The true order of the pieces
 * is not known and their order in this sequence record is

```

misc_feature 143035..147999
                /note="assembly_fragment"
misc_feature 148100..151734
                /note="assembly_fragment"
misc_feature 151835..155576
                /note="assembly_fragment"
misc_feature 155677..159193
                /note="assembly_fragment"
misc_feature 159294..161359
                /note="assembly_fragment"
misc_feature 161460..163515
                /note="assembly_fragment clone_end:r7 vector_side:right"
misc_feature 163616..166614
                /note="assembly_fragment clone_end:SP6 vector_side:left"
misc_feature 166715..169056
                /note="assembly_fragment"
misc_feature 169157..171527
                /note="assembly_fragment"
misc_feature 171628..173248
                /note="assembly_fragment"
misc_feature 173349..174815
                /note="assembly_fragment"
BASE COUNT 52310 a 30923 c 32183 g 57399 t 2000 others
ORIGIN

```

```

Query Match      64.0%; Score 16; DB 82; Length 174815;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 10 gcaagtcacagaattta 25
Db 44384 GCAGTCACAGATTTA 44369

```

```

RESULT 21
AC020734/c
LOCUS
DEFINITION HOMO SAPIENS 196172 bp DNA HTG 07-JUL-2000
SEQUENCE, 26 unordered pieces.
AC020734
AC020734.3 GI:8570370
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 196172)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 196172)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (08-JAN-2000) Genome Sequencing Center, Washington
REFERENCE University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL MO 63108, USA
COMMENT On Jun 17, 2000 this sequence version replaced gi:7024167.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH065114
----- Summary Statistics -----
Sequencing vector: M13; 85%
Sequencing vector: plasmid; 15%
Chemistry: Dye-primer ET; 85% of reads
Chemistry: Dye-terminator Big Dye; 15% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 184493 bases at least Q40
Consensus quality: 187694 bases at least Q30

```

```

Consensus quality: 189523 bases at least Q20
Insert size: 164000; agarose-fp
Insert size: 193672; sum-of-ctrls
Quality coverage: 4.86 in Q20 bases; agarose-fp
Quality coverage: 4.16 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 1451: contig of 1451 bp in length
* 1452: gap of unknown length
* 1552: contig of 2042 bp in length
* 3593: gap of unknown length
* 3594: gap of unknown length
* 3694: 7120: contig of 3427 bp in length
* 7121: 7220: gap of unknown length
* 7221: 9653: contig of 2433 bp in length
* 9654: 9753: gap of unknown length
* 9754: 12670: contig of 2917 bp in length
* 12671: 12770: gap of unknown length
* 12771: 13795: contig of 3025 bp in length
* 13796: 15895: gap of unknown length
* 15896: 17819: contig of 1924 bp in length
* 17820: 17919: gap of unknown length
* 17920: 20983: contig of 3064 bp in length
* 20984: 21083: gap of unknown length
* 21084: 25206: contig of 4123 bp in length
* 25207: 25306: gap of unknown length
* 25307: 28986: contig of 3680 bp in length
* 28987: 29086: gap of unknown length
* 29087: 32554: contig of 3468 bp in length
* 32555: 32654: gap of unknown length
* 32655: 37664: contig of 4810 bp in length
* 37665: 37664: gap of unknown length
* 37665: 42541: contig of 4977 bp in length
* 42542: 42641: gap of unknown length
* 42642: 47417: contig of 4776 bp in length
* 47418: 47517: gap of unknown length
* 47518: 52947: contig of 5430 bp in length
* 52948: 53047: gap of unknown length
* 53048: 58091: contig of 5044 bp in length
* 58092: 58191: gap of unknown length
* 58192: 65147: contig of 6956 bp in length
* 65148: 65247: gap of unknown length
* 65248: 71454: contig of 6207 bp in length
* 71455: 71554: gap of unknown length
* 71555: 76605: contig of 8051 bp in length
* 76606: 79705: gap of unknown length
* 79706: 88670: contig of 8965 bp in length
* 88671: 88770: gap of unknown length
* 88771: 97995: contig of 9225 bp in length
* 97996: 98095: gap of unknown length
* 98096: 109354: contig of 11259 bp in length
* 109355: 109454: gap of unknown length
* 109455: 122284: contig of 12830 bp in length
* 122285: 122384: gap of unknown length
* 122385: 139948: contig of 17564 bp in length
* 139949: 140048: gap of unknown length
* 140049: 165725: contig of 25677 bp in length
* 165726: 165825: gap of unknown length
* 165826: 196172: contig of 30347 bp in length.

```

FEATURES

```

source
misc_feature
1..196172
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/clone="RP11-665114"
1..1451
/note="assembly_name:Contig4"

```

```
misc_feature 1552. .3593 /note="assembly_name:Contig5"  
misc_feature 3694. .7120 /note="assembly_name:Contig6"  
misc_feature 7221. .9653 /note="assembly_name:Contig7"  
misc_feature 9754. .12670 /note="assembly_name:Contig8"  
misc_feature 12771. .15795 /note="assembly_name:Contig9"  
misc_feature 15896. .17819 /note="assembly_name:Contig10"  
misc_feature 17920. .20983 /note="assembly_name:Contig11"  
misc_feature 21084. .25206 /note="assembly_name:Contig12"  
misc_feature 25307. .28986 /note="assembly_name:Contig13"  
misc_feature 29087. .32554 /note="assembly_name:Contig14"  
misc_feature 32655. .37464 /note="assembly_name:Contig15"  
misc_feature 37565. .42541 /note="assembly_name:Contig16"  
misc_feature 42642. .47417 /note="assembly_name:Contig17"  
misc_feature 47518. .52947 /note="assembly_name:Contig18"  
misc_feature 53048. .58091 /note="assembly_name:Contig19"  
misc_feature 58192. .65147 /note="assembly_name:Contig20"  
misc_feature 65248. .71454 /note="assembly_name:Contig21"  
misc_feature 71555. .79605 /note="assembly_name:Contig22"  
misc_feature 79706. .88670 /note="assembly_name:Contig23"  
misc_feature 88771. .97995 /note="assembly_name:Contig24"  
misc_feature 98096. .109354 /note="assembly_name:Contig25"  
clone_end:T7  
vector_side:right"  
misc_feature 109455. .122384 /note="assembly_name:Contig26"  
misc_feature 122385. .139948 /note="assembly_name:Contig27"  
misc_feature 140049. .165725 /note="assembly_name:Contig28"  
misc_feature 165826. .196172 /note="assembly_name:Contig29"  
BASE COUNT 64027 a 34406 c 33812 g 61398 t 2529 others  
ORIGIN
```

Query Match	Score 16;	DB 66;	Length 161172;
Best Local Similarity	100.0%;	Pred. No. 4.7;	
Matches 16;	Conservative 0;	Mismatches 0;	Indels 0;
			Gaps 0;
OY	10	gcaagtcacagaatla	25
Db	94789	CCAACTCACAGAAATTA	94774
RESULT	22		
LOCUS	AR080743		
DEFINITION	Sequence 48 from patent US 5968826.		
ACCESSION	AR080743		
VERSION	AR080743.1	GI:10007473	
KEYWORDS			
SOURCE	Unknown.		

ORGANISM	unknown.
	unclassified.
REFERENCE	1 (bases 1 to 1771)
AUTHORS	Bennett,C.Frank,Comdon,T.P. and Cowser,L.M.
TITLE	Antisense inhibition of integrin .alpha.4 expression
JOURNAL	Patent: US 5968836-A 48 19-OCT-1999;
FEATURES	location/Qualifiers
source	1..1771
	/organism="unknown"
BASE COUNT	371 a 471 c 519 g 410 t
ORIGIN	

Query Match	60.0%	Score 15	DB 9	Length 1771
Best Local Similarity	100.0%	Pred. No. 37		
Matches 15, Conservative	0	Mismatches	0	Indels 0; Gaps 0;
QY	10	gcaagtcacagaaatt	24	
Db	312	GCACAGTCACAGAAATT	326	

RESULT	23
MUSINTESUB	
LOCUS	MUSINTESUB 1771 bp DNA
DEFINITION	Mus musculus (clone A4G16) integrin alpha-4 subunit gene, 5' end cds.
ACCESSION	L20788
VERSION	L20788.1 GI:309416
KEYWORDS	integrin alpha-4 subunit.
SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE	1 (bases 1 to 1771)
AUTHORS	De Meirman, C., Scholten, E., Jaspers, M., Ongena, K., Mathijs, G., Maynen, P., and Cassiman, J. J.
TITLE	Cloning and characterization of the promoter region of the murine alpha 4 integrin subunit
JOURNAL	DNA Cell Biol. 13 (7), 743-754 (1994)
MEDLINE	95290094

FEATURES	Location/Qualifiers
source	1. .1771

```
GC_signal      /organelm="Mus_musculus"  
                /db_xref="taxon:10090"  
                /cell_line="3T3"  
                /cell_type="Fibroblasts"  
                /tissue_lib="NIH 3T3 in lambda-FIXII"  
                503..508  
                /note="putative"  
GC_signal      845..850  
GC_signal      952..1389  
exon            complement(1055..1060)  
GC_signal      john(1193..1389,1711..1771)  
gene            /gene="VLA-4"  
                john(1193..1389,1711..>1771)  
DSS
```

	371 a	471 c	519 g	410 t
intron	1390.	.1710		
exon	1711.	.1771		
	/gene="VLA-4"			
BASE COUNT				
ORIGIN				

Query Match	60.08;	Score 15;	DB 94;	Length 1771;
Best Local Similarity	100.08;	Pred. No. 37;		

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacagaatt 24
 |||
 Db 312 GCAAGTCACAGAAATT 326

RESULT 24
 LOCUS G07181 1829 bp DNA STS 19-OCT-1995
 DEFINITION human STS WI-9211.
 ACCESSION G07181
 VERSION G07181.1 GI:860426
 KEYWORDS STS sequence; primer; sequence tagged site.
 SOURCE human STS derived from sequences in dbEST and the Unigene collection.

ORGANISM

Homo sapiens
 Eukaryota; Mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata;
 Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
 Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 1829)

REFERENCE

1 Hudson, T.
 Whitehead Institute/MIT Center for Genome Research; Physically
 mapped ESTs
 Unpublished (1995)

COMMENT

Contact: Thomas Hudson
 Whitehead Institute/MIT Center for Genome Research
 Whitehead Institute for Biomedical Research
 9 Cambridge Center, Cambridge MA 02142 USA
 Tel: 617 252 1900
 Fax: 617 252 1902
 Email: thudson@genome.wi.mit.edu

Primer A: CAGCCACTACTACATCGCATTCAGC

Primer B: GACAGAAATGATGATATTAATGCG

STS size: 111

PCR Profile:

Presoak:

Denaturation:

Annealing: 56 degrees C

Polymerization:

PCR Cycles: 35

Thermal Cycler:

Protocol:

Template: 10 ng

Primer: each 5 pM

dNTPs: each 4 mM

Tag Polymerase: 0.025 units/ul

Total Vol: 20 ul

Buffer:

MgCl2: 1.5 mM

KCl: 50 mM

Tris-HCL: 10 mM

pH: 9.3

FEATURES

Prepared with primer pairs derived from D13644 -- Unigene.

Location/Qualifiers

1..1829

/organism="Homo sapiens"

STS

Primer_bind

891..912

complement(977..1001)

BASE COUNT

539 a 321 c 320 g 590 t 59 others

ORIGIN

Query Match 60.0%; Score 15; DB 54; Length 1829;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaatta 25
 |||
 Db 1689 CAACTCACAGAAATTA 1675

RESULT 25
 LOCUS MMINTEGR01 1949 bp DNA ROD 02-OCT-1996
 DEFINITION Mus musculus alpha-4 integrin gene, exons 1 and 2.
 ACCESSION U34627
 VERSION U34627.1 GI:1173582
 KEYWORDS 1 of 22
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (sites)
 De Meirsmen, C., Scholten, E., Jaspers, M., Ongena, K., Matthijs, G.,
 Marynen, P., and Cassiman, J.J.
 Cloning and characterization of the promoter region of the murine
 alpha-4 integrin subunit
 DNA Cell Biol. 13 (7), 743-754 (1994)
 95290094
 2 (bases 1 to 1949)
 De Meirsmen, C., Jaspers, M., Scholten, E., and Cassiman, J.J.
 The genomic structure of the murine alpha 4 integrin gene
 DNA Cell Biol. 15 (7), 595-603 (1996)
 96326295
 3 (bases 1 to 1949)
 De Meirsmen, C.
 Direct Submission
 Submitted (23-AUG-1995) Catherine De Meirsmen, Center for Human
 Genetics, University of Leuven, Herestraat 49, Leuven 3000, Belgium

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

Query Match 60.0%; Score 15; DB 94; Length 1949;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacagaatt 24
 |||
 Db 312 GCAAGTCACAGAAATT 326

RESULT 26

LOCUS ECHK311M 2340 bp DNA BCT 25-SEP-1998

DEFINITION E. coli HK311M gene.

ACCESSION X82231

VERSION X82231.1 GI:3660492

KEYWORDS alpha-polypeptide; beta-polypeptide; R.EcoHK311 protein.

SOURCE Escherichia coli.

ORGANISM Escherichia coli.

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

Escherichia.

1 (bases 1 to 2340)

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

AUTHORS Lee, K.F., Kam, K.M. and Shaw, P.C.
 TITLE A bacterial methyltransferase M.EcoHK311 requires two proteins for
 JOURNAL in vitro methylation
 MEDLINE Nucleic Acids Res. 23 (1), 103-108 (1995)
 REFERENCE 95175351
 AUTHORS 2 (bases 1 to 2340)
 TITLE Shaw, P.
 JOURNAL Direct Submission
 REMARK Submitted (18-OCT-1994) P. Shaw, Dept. of Biochemistry, The Chinese
 JOURNAL University of Hong Kong, Shatin NT, HONG KONG
 TITLE revised by [3]
 REFERENCE 3 (bases 1 to 2340)
 AUTHORS Shaw, P.
 JOURNAL Direct Submission
 TITLE Submitted (16-DEC-1997) P. Shaw, Dept. of Biochemistry, The Chinese
 JOURNAL University of Hong Kong, Shatin NT, HONG KONG
 REFERENCE 4 (bases 1 to 2340)
 TITLE Lee, K.F., Shaw, P.C., Picone, S.J., Wilson, G.G. and Lunnen, K.D.
 JOURNAL Sequence comparison of the EcoHK311 and EaeI
 TITLE restriction-modification systems suggests an intergenic transfer of
 genetic material
 JOURNAL Biol. Chem. 379 (4-5), 437-441 (1998)
 MEDLINE 98290299
 COMMENT On Sep 28, 1998 this sequence version replaced gi:639971.
 FEATURES
 SOURCE
 1. 2340
 Location/Qualifiers
 /organism="Escherichia coli"
 /strain="HK31"
 /db_xref="taxon:562"
 407..1336
 /gene="ecoHK311m alpha"
 407..1336
 /gene="ecoHK311m beta"
 /function="methyltransferase"
 /codon_start=1
 /transl_table=11
 /product="M.EcoHK311 alpha polypeptide"
 /protein_id="CA57706.1"
 /db_xref="GI:3660493"
 /db_xref="SPTREMBL:087740"
 /translation="MKKKPLKQYKVSFFAGIGGPDIGLEKAGMEVVFQCEINKEQCK
 VERKNSKYLHTDITRLNADIPESNWCQGGPCDDVSSANQGRKGLGARSGLFY
 TVAKLTERRPEMLITENYVGLINSHGDDPKVDTLVEFGVSMRWLDARFETP
 ORRRRYIVASLIDMSARVLPFCATRTVDQGGGERNAGASATGTSKADIVSIO
 HASIGRKATAGPQAKGYRNDGETYTTIDSKSSDVAAPDAFRVRASTGISLEDSNR
 FRAVGNAAVAVPVEWLGKRIVLVDQVINEALS"
 717..1247
 /gene="ecoHK311m beta"
 717..1247
 /gene="ecoHK311m beta"
 /function="methyltransferase"
 /codon_start=1
 /transl_table=11
 /product="M.EcoHK311 polypeptide beta"
 /protein_id="CA57705.1"
 /db_xref="GI:3660494"
 /db_xref="SPTREMBL:047257"
 /translation="MONSSKRESJNGILLKMPDOSTATMDKTSKISSIRWSNCGAF
 RGEVWQNTLEHPSVEECTLSQVLETCAPLESFLNPEQSLINRAKRGOMLPEPL
 LOAYOKOISLISNOVLEDERQPODLKDKGTMEKPTLLTQEEVQMLVYVRMLPSEY
 RLOGFPENWTLIDSEC"
 complement(1317..2273)
 /gene="ecoHK311r"
 complement(1317..2273)
 /gene="ecoHK311r"
 /function="restriction endonuclease"
 /codon_start=1
 /transl_table=11
 /product="R.EcoHK311 protein"
 /protein_id="CA57707.1"
 /db_xref="GI:3660495"
 /db_xref="SPTREMBL:087741"
 /translation="MAVSIAYELQRTVDQDTIVPAKGTQIOCPFMDRTCDKASKVK
 CDS
 gene
 CDS
 gene
 CDS
 gene

Query Match
 Best Local Similarity 100.0%; Score 15; DB 2; Length 2340;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 acattgtgcgaagtc 16
 Db 906 ACATTGTCGCAAGTC 920
 RESULT 27
 LOCUS AX098224/C 3748 bp DNA PAT 30-MAR-2001
 DEFINITION Sequence 136 from Patent WO0118542.
 ACCESSION AX098224
 VERSION AX098224.1 GI:13515348
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 3748)
 AUTHORS Lee, J., Thomsho, P. and Lillie, J.
 TITLE Identification, assessment, prevention, and therapy of ovarian
 JOURNAL cancer. WO 0118542-A 136 15-MAR-2001;
 JOURNAL Millennium Predictive Medicine, Inc. (US)
 FEATURES
 SOURCE
 1. 3748
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 1110 a 825 c 761 g 1052 t
 ORIGIN
 Query Match
 Best Local Similarity 100.0%; Score 15; DB 10; Length 3748;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 caatgcacagaatta 25
 Db 2627 CAAGTCACAGAAATTA 2613
 RESULT 28
 LOCUS D13644 4602 bp mRNA PRI 10-JUL-1997
 DEFINITION Human mRNA for KIAA0019 gene, complete cds.
 ACCESSION D13644
 VERSION D13644.1 GI:1531551
 KEYWORDS KIAA0019.
 SOURCE Homo sapiens male myeloblast cell line KG-1 cDNA to mRNA.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 4602)
 AUTHORS Nomura, N.
 TITLE Direct Submission
 JOURNAL Submitted (11-NOV-1992) to the DDBJ/EMBL/GenBank databases. Nobuo
 Nomura, Kazusa DNA Research Institute, Gene Structure 1; 1532-3
 yana, Kisarazu, Chiba 292, Japan (E-mail:cdna1nfo@kazusa.or.jp,
 URL:http://www.kazusa.or.jp, Tel:0438-52-3930, Fax:0438-52-3931)
 REFERENCE 2 (bases 1 to 4602)
 AUTHORS Nomura, N., Miyajima, N., Kawarabayashi, Y. and Tabata, S.
 TITLE Prediction of new human genes by entire sequencing of randomly

JOURNAL sampled cDNA clones
 REFERENCE Unpublished (1994)
 AUTHORS Nomura,N., Miyajima,N., Sazuka,T., Tanaka,A., Kawarabayashi,Y., Sato,S., Nagase,T., Seki,N., Ishikawa,K. and Tabata,S.
 TITLE Prediction of the coding sequences of unidentified human genes. I. The coding sequences of 40 new genes (K1AA0001-K1AA0040) deduced by analysis of randomly sampled cDNA clones from human immature myeloid cell line KG-1
 JOURNAL DNA Res. 1 (1), 27-35 (1994)
 MEDLINE 96051387
 REMARK Erratum: [[published erratum appears in DNA Res 1995 Aug 31;2(4):211]]
 REFERENCE 4 (sites)
 AUTHORS Nomura,N., Miyajima,N., Sazuka,T., Tanaka,A., Kawarabayashi,Y., Sato,S., Nagase,T., Seki,N., Ishikawa,K. and Tabata,S.
 TITLE Prediction of the coding sequences of unidentified human genes. I. The coding sequences of 40 new genes (K1AA0001-K1AA0040) deduced by analysis of randomly sampled cDNA clones from human immature myeloid cell line KG-1 (supplement)
 JOURNAL DNA Res. 1 (1), 47-56 (1994)
 MEDLINE 96051389
 REFERENCE 5 (bases 1 to 4602)
 AUTHORS Matoskova,B., Wong,W.T., Seki,N., Nagase,T., Nomura,N., Robbins,K.C. and Di Fiore,P.P.
 TITLE RN-tre identifies a family of tre-related proteins displaying a novel potential protein binding domain
 JOURNAL Oncogene 12 (12), 2563-2571 (1996)
 COMMENT On Sep 10, 1996 this sequence version replaced gi:286016.
 FEATURES
 source
 1..4602
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_line="KG-1"
 /cell_type="myoblast"
 /chromosome="10"
 /map="10p13"
 /sex="male"
 <1..279
 280..2766
 /gene="K1AA0019"
 280..2766
 /gene="K1AA0019"
 /codon_start=1
 /product="protein related N-terminus of tre oncogene"
 /protein_id="BA02807.1"
 /db_xref="GI:2104571"
 /translation="MNSDDVALKIAQERAEIVAKYDREGAEIEPWEDADIVYKKV LDRFGLEHEELPDHNAVEROKHEIERTTKMLKMGEMKYKTEKPHRRITKGIPLQIRGEVWALLLEIPKMKETRLDYSLKHAAGCSPDRIOLDLVNPFRRHIMRD RYGVQOOSLFHVAASYINTEVGYCGOSQITALLMYNEDDAFWALVKIFSPGRK AHGFFVQGFPKLRFQEHHEKLINKELKIKOHILDSOETISFYTKMKPFQCDLRT PPTLNLRIMDIYEEGERVLTAMSTTIILKHKMLKLSMEELVEFQETLAKDFEEDDVITQLOISMTELKRAKLDLDEPEKEDYPRKPLGULPELQSWGVIHLNNGORSV GRPSPLASGRREGAPRRHREHSPIPOSTGTPERRAPRRKSVEESKLLKDEADFO RKILPSGDDSSROYNHAANONSATSNIRKEFVPRKMSPDVSATERTAKYTMGCK RAHPALAVTVPGPAEVRVSNVPRKALDAEGKRGSTASQYDNPAGPELNSGASVE EALERAYSOSPRAHYLPSPRRKHAEPSSSPKSVNKFYKVPSPSHARPQSLDDEAR GLAHPYSNPPIVYHNGNSPKHPHTANSSSPASQPSQGTOLNRRPHSGTLLSVSEPS KYSRSPILVBSRILEYLPVTGAGGISGNSGSPKNGKLLIPVDYLDPDNTWSEVS YTRPETQOSQSMTRDSRGNLPRKYSFQIAPDQGLPAVSDSPVRKASPAADAS PSQYPSGPPPAHYHNRNDGLSIQESVLL"
 2767..4602
 3'UTR
 BASE COUNT 1340 a 1041 c 999 g 1222 t
 ORIGIN
 Query Match 60.0%; Score 15; DB 91; Length 4602;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 caagtcacagagatta 25 .

Db 4442 CAAGTCACAGAAATTA 4448
 RESULT 29
 AF039582
 LOCUS
 DEFINITION
 AF039582 4636 bp DNA BCT 05-DEC-2000
 Enterobacter aerogenes putative integrase (int(p)) gene, partial cds; and EaeI methyltransferase alpha subunit (eaeIM-a); EaeI methyltransferase beta subunit (eaeIM-b); and EaeI restriction endonuclease (eaeCR) genes, complete cds.
 AF039582
 AF039582.1 GI:2760953
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Enterobacter aerogenes.
 Enterobacter aerogenes.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Enterobacter.
 REFERENCE
 1 (bases 1 to 4636)
 Lee,K.F., Shaw,P.C., Picone,S.J., Wilson,G.G. and Lunnen,K.D.
 TITLE Sequence comparison of the EcoHK311 and EaeI restriction-modification systems suggests an intergenic transfer of genetic material
 JOURNAL Biol. Chem. 379 (4-5), 437-441 (1998)
 MEDLINE 98290299
 PUBMED 9628335
 JOURNAL 2 (bases 1 to 4636)
 REFERENCE
 2 (bases 1 to 4636)
 Lunnen,K.D., Picone,S.J. and Wilson,G.G.
 TITLE Direct Submission
 JOURNAL Submitted (22-DEC-1997) Research, New England Biolabs, 32 Tozer Road, Beverly, MA 01915, USA
 FEATURES
 source
 1..4636
 /organism="Enterobacter aerogenes"
 /db_xref="taxon:548"
 <1..1035
 /gene="int(p)"
 <1..1035
 /gene="int(p)"
 /note="similar to phage p4 integrase SP:p39347"
 1..1035
 /gene="int(p)"
 /note="similar to phage p4 integrase SP:p39347"
 /codon_start=1
 /transl_table=11
 /product="putative integrase"
 /protein_id="AAB95339.1"
 /db_xref="GI:2760957"
 /translation="K1VAAGLDPREHRAVKEEQTGCATTFESVAREMHANKKTEE HSRVLSLEIDNLEPAIGKRSIDFSFTRDLVPIKVEATGRLEVASRLQORTMAIR YAVOSGLIDYNDPAQEMAGAVASNRVRRPALERKLELRLORIDSYGRPLTRAVLT LILFIRSELRFAPWPEIDFERAMWTIPAREALIEGVKISORSGSKRPTLVLSRQ ALLEIKOYHKSGRDPYFGVDHNPDRKPMSENIVYKALRWAGYDTKYEVGCHGRTTA CSSLSLSNRRDAVEROMSHMERNVRAATIIKAEHLDERKLMLOWADFLDANROK GISPDTAKA"
 1228..2157
 /gene="eaeIM-a"
 1228..2157
 /gene="eaeIM-a"
 /note="W.aeIIa"
 /function="converts YGGCCR to YGGMCCR"
 /codon_start=1
 /transl_table=11
 /product="EaeI methyltransferase alpha subunit"
 /protein_id="AAB95336.1"
 /db_xref="GI:2760954"
 /translation="MKKKPLKQYKVSFFAGTGGFDLGEKAGMEVYFQCEINKFCOK VLRKNWNPVPLHDTITQLNADLIPSSVWVGFPDQVSSANOGKRRKREGASGLFET TYAKLIERKPEWLVLENVGLNLSHGQPKFVITPLVELIGVSMRVADAYFGTPT QRRRYIVTASLIGDMSARVLFEGATRIYIDKOGQGERTAASAGSDTKSKADITYSIO HASIGRKATAGPOAKGYRNDGERTYLLDSRSSAVCAVPAADAFVRASAGISRLDSNR FRAVGNVAVPIYEWICKRIVLDQGMNALS"
 1538..2068
 /gene="eaeIM-b"
 1538..2068
 /gene="eaeIM-b"
 /gene="eaeIM-b"

```

/note="M.EaeIb: translated from 3'-half of the eaeIb-a
gene in an alternative reading frame"
/codon_start=1
/transl_table=11
/product="EaeI methyltransferase beta subunit"
/db_xref="GI:2760955"
/translation="MONSSKKENLNGLLKMPDCCSTATMDKTSKISSIRWSNGMAF
RGEYMMONTLEHPSVEECTLSQVLETCAPLESFLNPEQLESLINAKRGOMLPEPL
IOAYKOISILSSMOVLDEKOPDLKQKDTGTTEKTHLQEEVOMLYVRMLPSBYE
RLOGEPENMTLIDSEQ"
/complement(2138..3085)
/gene="eaeIb"
/complement(2138..3085)
/function="cleaves between Y and G in YGGCCR"
/note="R.EaeI: unable to cleave methylated ds sequence
YGGCCR"
/codon_start=1
/transl_table=11
/product="EaeI restriction endonuclease"
/protein_id="AAB95338.1"
/db_xref="GI:2760956"
/translation="MAVSTAEYRGORTDVOOIVPVSKGTIAQCPMDRTPCKVSKNN
HPVGSYRKPDPSTIWCCEHRLCTSRKKTIVSGEQVDNIVLHQHRLKRVAKLI
YQPELDSEITGVKREVNIPLPDSNTHADYVKNFSGRGVDEVLEMOGGGETSS
TGAITRHAAMADLEPPNLEILRPVAANTETNAMRQEQFLVGNVDDTGKIV
FVAGSLYDLHRRFPNANLRLKHNMTLCIAFKEDTSDPEQPCPIPLIIDSRTLL
FTWSTFVRLTQOGAPRPELFGSGFLRLDNSTVITMPR"
BASE COUNT      1362 a      893 c      1074 g      1307 t
ORIGIN

```

```

Query Match      60.0%; Score 15; DB 1; Length 4636;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      2 acattgtcgcaagtc 16
      |||||
Db      1727 ACATTGTCGCAAGTC 1741

```

```

RESULT 30
AF152313/c      5218 bp      mRNA      PRI      22-JUL-1999
LOCUS      Homo sapiens protocadherin alpha 5 (PCDH-alpha5) mRNA, complete
DEFINITION
ACCESSION      AF152313
VERSION
KEYWORDS
SOURCE
ORGANISM      human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS      Wu,Q. and Maniatis,T.
TITLE      A striking organization of a large family of human neural
cadherin-like cell adhesion genes
JOURNAL      Cell 97 (6), 779-790 (1999)
MEDLINE      99308636
PUBMED      10380929
REFERENCE
AUTHORS      Wu,Q. and Maniatis,T.
TITLE      Human protocadherin genes
JOURNAL      unpublished
REFERENCE
AUTHORS      Wu,Q. and Maniatis,T.
TITLE      Direct Submission
JOURNAL      Submitted (17-MAY-1999) Department of Molecular and Cellular
Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
02138, USA
FEATURES
source      1..5218
Location/Qualifiers

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q31"
/tissue-type="brain"
1..5218
/gene="PCDH-alpha5"
1..2811
/gene="PCDH-alpha5"
/function="cell-cell adhesion"
/codon_start=1
/product="protocadherin alpha 5"
/protein_id="AAD43707.1"
/db_xref="GI:5456914"
/translation="MYSRKSLSGRLLMLLLAYKAGSGOLHYSIPERAKGTFV
GRIADDLLELAELPRLEFVASKGKGLLEVNILNGILFVNSRIDRELCRRARECS
IHLVIVRPLQVPHVEVAVKDINDNPREFROQRULESRMPDSRPLEGASDLID
IGANQLRYRLNPNPEYFDLVKTEENFSELELYRSLDRETEQHRLLVAVDGGK
PELTGTVOLILNVLANDNAPEPDKSYVNLRLNAPSGLTVIKLNSADDEGINKEL
VYFESNLVLDVSKSKFLINSMTGELKVGSLDYEDYNSYEINIDAMKSTPPLSGHCK
VYKLLDNDNTPEKATITTLPLPYKEDAPLSVTVALISVSDRSGANQOVCSIMPHY
PKRLVSTFRKNTSLVLDASALDRESVSYELVVTARDOGSPLMATASVSEYAVVNN
APAFAPQYTVFVKENNPGGCHIFVSARDAQENALVSTLVERVGERPLSYVS
VHAESGKYVALQPLDHEVELOEQVSARDAGVPLGSNVTLQVFLDENDBALLV
PVGGTGAVSELVPSVYGAGHYAKVAVPDGQVSNMLSYELQAPASARIPFPG
LYTGRISTRSIDETEARRHRLVAVKHGPEPLTATVLSLVESGOARAKSSRAS
ACAVPEALVDVNYVYLIALCAVSSLLVITLLYTLARCAOPEACCTGKPTLLC
SSAVSWSTISQQRORVCSGEAPKPTIDMAFSPSLPGQPTSDNRPQNPWRKSASL
VPGNMSSVHLEAGILRAGGQPDQOWPVSAPPEPAGVSPVGVGAVNSWTR
YRGNPKOSGPELDPKFLIPGSPAILISIROEPTNSQIDKSDFTTFCKKETKKKKK
KKGNTOEKKEKGNSTDNSDQ"
BASE COUNT      1385 a      1207 c      1281 g      1345 t
ORIGIN

```

```

Query Match      60.0%; Score 15; DB 89; Length 5218;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      11 caegtccagaatta 25
      |||||
Db      4115 CAAGTCACAGAAATTA 4101

```

```

RESULT 31
AF152315/c      5221 bp      mRNA      PRI      22-JUL-1999
LOCUS      Homo sapiens protocadherin alpha 7 (PCDH-alpha7) mRNA, complete
DEFINITION
ACCESSION      AF152315
VERSION
KEYWORDS
SOURCE
ORGANISM      human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS      Wu,Q. and Maniatis,T.
TITLE      A striking organization of a large family of human neural
cadherin-like cell adhesion genes
JOURNAL      Cell 97 (6), 779-790 (1999)
MEDLINE      99308636
PUBMED      10380929
REFERENCE
AUTHORS      Wu,Q. and Maniatis,T.
TITLE      Human protocadherin genes
JOURNAL      unpublished
REFERENCE
AUTHORS      Wu,Q. and Maniatis,T.
TITLE      Direct Submission
JOURNAL      Submitted (17-MAY-1999) Department of Molecular and Cellular
Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
02138, USA

```

FEATURES
source
Location/Qualifiers
1..5221
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q31"
/tissue_type="brain"
1..5221
/gene="PCDH-alpha7"
1..2814
/gene="PCDH-alpha7"
/function="cell-cell adhesion"
/codon_start=1
/product="protocadherin alpha 7"
/protein_id="A043709.1"
/db_xref="GI:5456918"
/translation="MPCPNGYDPGGRHLLFLIILAMWEGSGQLHYVPEAKHGF
VGRADQGLLELAEVPLRFVAVCKFRGDLLEVNQILFVNSRIDREELGKRSAC
SIHLEIVLPLQVHVDVVDKINDNPVFEREDQKVSSEAPLDSHFLLEGASDA
DIGENALTYRLSENEYFLDPTSNOCVKPLGLVLRKLDRETPELHLLITATDGG
KPELTGTVOLLITVLDNNDAVPEDRTLYTKLIPENNSIGTLVHRNASDLDELND
ITVPEVYVNDNAQQLTSLSLPDPDQGTITLISVDFRGVNGVYCSLTPR
VPEKLVSTFKNYSLVLDSDALDRSVAYELVYTRADGSPSLMTATASVEYADVND
NAPAEADPEYVAVKNNPGCHITFTVSAGDADQKNAIVSYSLVELRGESALSYV
SVHESGKVYALQPLDHELELLQVSAKDADGVPPLGNSVTLQGVLDNDAPALL
APRVGGGAVRELVPVSGVAVAVAYADAGYNAMLSYELQVPAVASIPRV
GLYGETSTRALDETDAPRHLIVLWKDGPSTPATATVSLVSGCAPKASSRA
SLGTAGPETELVDVNYLITACVSSLLITLTLTALYALKSAPSSGACSLKPTLV
CSSAVGSKNSQQRKRVCSGEGPCKDLMFSPSLPQPSSTDPNPQMDKYSV
LRAGMHSVHLEAGILRAGGPGDQMPVYSSATPEPEAGEVSPVAGAVNSWTF
KYGNGNSKSGELPDKFLLPGSPALISIROEPTNSQIDKSPFITPGKKEETKKKK
KKKKNKTOEKKEKNGSTTNSDQ"

BASE COUNT 1333 a 1242 c 1295 g 1351 t
ORIGIN

Query Match 60.0%; Score 15; DB 89; Length 5221;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caagtcacagaatta 25
|||||

Db 4118 CAAGTCACAGAATTA 4104

RESULT 32
AF152308 5233 bp mRNA PRI 22-JUL-1999
LOCUS AF152308
DEFINITION Homo sapiens protocadherin alpha 12 (PCDH-alpha12) mRNA, complete
cds.
ACCESSION AF152308
VERSION AF152308.1 GI:5456901
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 5233)
AUTHORS Wu,Q. and Maniatis,T.
TITLE A striking organization of a large family of human neural
cadherin-like cell adhesion genes
JOURNAL Cell 97 (6), 779-790 (1999)
MEDLINE 99308636
PUBMED 10380929
REFERENCE 2 (bases 1 to 5233)
AUTHORS Wu,Q. and Maniatis,T.
TITLE Human protocadherin genes
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 5233)
AUTHORS Wu,Q. and Maniatis,T.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-1999) Department of Molecular and Cellular

FEATURES
source
Location/Qualifiers
02138, USA
1..5233
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q31"
/tissue_type="brain"
1..5233
/gene="PCDH-alpha12"
1..2826
/gene="PCDH-alpha12"
/function="cell-cell adhesion"
/codon_start=1
/product="protocadherin alpha 12"
/protein_id="A043702.1"
/db_xref="GI:5456902"
/translation="MTITPRGSGRRLLSLILAMFVSGQLHYVPEAKHGF
VGRADQGLLELAEVPLRFVAVSKRHGDLLEVNQILFVNSRIDREELGKRSAC
SIHLEIVLPLQVHVDVVDKINDNPVFEREDQKVSSEAPLDSHFLLEGASDA
DIGENALTYRLSENEYFLDPTSNOCVKPLGLVLRKLDRETPELHLLITATDGG
KPELTGTVOLLITVLDNNDAVPEDRTLYTKLIPENNSIGTLVHRNASDLDELND
ITVPEVYVNDNAQQLTSLSLPDPDQGTITLISVDFRGVNGVYCSLTPR
VPEKLVSTFKNYSLVLDSDALDRSVAYELVYTRADGSPSLMTATASVEYADVND
NAPAEADPEYVAVKNNPGCHITFTVSAGDADQKNAIVSYSLVELRGESALSYV
SVHESGKVYALQPLDHELELLQVSAKDADGVPPLGNSVTLQGVLDNDAPALL
APRVGGGAVRELVPVSGVAVAVAYADAGYNAMLSYELQVPAVASIPRV
GLYGETSTRALDETDAPRHLIVLWKDGPSTPATATVSLVSGCAPKASSRA
SLGTAGPETELVDVNYLITACVSSLLITLTLTALYALKSAPSSGACSLKPTLV
CSSAVGSKNSQQRKRVCSGEGPCKDLMFSPSLPQPSSTDPNPQMDKYSV
LRAGMHSVHLEAGILRAGGPGDQMPVYSSATPEPEAGEVSPVAGAVNSWTF
KYGNGNSKSGELPDKFLLPGSPALISIROEPTNSQIDKSPFITPGKKEETK
KKKKNKTOEKKEKNGSTTNSDQ"

BASE COUNT 1364 a 1213 c 1305 g 1351 t
ORIGIN

Query Match 60.0%; Score 15; DB 89; Length 5233;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caagtcacagaatta 25
|||||

Db 4130 CAAGTCACAGAATTA 4116

RESULT 33
AF152312 5251 bp mRNA PRI 22-JUL-1999
LOCUS AF152312/c
DEFINITION Homo sapiens protocadherin alpha 4 (PCDH-alpha4) mRNA, complete
cds.
ACCESSION AF152312
VERSION AF152312.1 GI:5456911
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 5251)
AUTHORS Wu,Q. and Maniatis,T.
TITLE A striking organization of a large family of human neural
cadherin-like cell adhesion genes
JOURNAL Cell 97 (6), 779-790 (1999)
MEDLINE 99308636
PUBMED 10380929
REFERENCE 2 (bases 1 to 5251)
AUTHORS Wu,Q. and Maniatis,T.
TITLE Human protocadherin genes
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 5251)
AUTHORS Wu,Q. and Maniatis,T.

TITLE Direct Submission
JOURNAL Submitted (17-MAY-1999) Department of Molecular and Cellular
Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
02138, USA

FEATURES

source

Location/Qualifiers
1..5251

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="5"

/map="5q31"

/tissue-type="brain"

1..5251

/gene="PCDH-alpha4"

1..2844

/gene="PCDH-alpha4"

/function="cell-cell adhesion"

/product="protocadherin alpha 4"

/protein_id="AA043706.1"

/db_xref="GI:5456912"

/translation="MEFSGSGSRRLLLLLLLAWEANGQULHYVSEAKHCTE
VGRADQGLLELAEVLPRLFRVASKRGGLLEVNQNGILFVNSRIDRELCRSAC
SHELVYDPLQVHFVDEVDINDNPVFPATOKLISASRPDLSRFPLEGASDA
DIGNALTYRLSPNEYFLEKPPDEIKKGLILKSLDREAEIFLVATDGG
KPELCTGQVLLTVLDANNAFDPDTIKVRLLEVPNGCTVIKINAQDLGNGD
IVTSFNDISPNVSKFHIDPTGQITVIGYIDFEBSKTEIIVGIDKQQLPLSGHC
RVIVEVEDNDNDVPLEFSLPIREDAPLGVIALISVSDKMGVNLVTCSTLSD
VPEKLSTFENYSLVLDALDRESAYELVVTADGSPSLMATAVSVEVADND
NAPAFQETVYVKKENPPGCHIFVSAAMDADQNALVSLVERVGERALSSYV
SVAESGKYVALQPLDHELELLOFVOTADACVPLGNSVTLQVFLDENAPALL
APRAGTGAVSELVPSVGVGHVAKVRVADDSGNMLSYELQPGGAGATIPRV
GLYTGELSTRALDETDPARHLVLYKHGEPAULATATVLSYVESCQAPASSRA
LVGAVGDAALVDVNYLITATCAVSSLLVLLLTALKCSALPTGCAKCPKPLV
CSSAVGMSYQQQRQVSCGEGPPTDLMAFSPSLPSDRDQDTQTTEFSAKPRQ
PNPDMRYASLRAGMHSVHLEAGILRAGGPGDQMPVSSATPEPEGEVSPVQ
AGVNSNSWTFYFGNPKOSGCELPKFTIPGSPALISIROEPTNSQIDKSDFTFG
GKEETKKKKKKKGNKQEKKEKGNSTTDSQ"

gene
CDS

1..2844

/gene="PCDH-alpha4"

/function="cell-cell adhesion"

/product="protocadherin alpha 4"

/protein_id="AA043706.1"

/db_xref="GI:5456912"

/translation="MEFSGSGSRRLLLLLLLAWEANGQULHYVSEAKHCTE
VGRADQGLLELAEVLPRLFRVASKRGGLLEVNQNGILFVNSRIDRELCRSAC
SHELVYDPLQVHFVDEVDINDNPVFPATOKLISASRPDLSRFPLEGASDA
DIGNALTYRLSPNEYFLEKPPDEIKKGLILKSLDREAEIFLVATDGG
KPELCTGQVLLTVLDANNAFDPDTIKVRLLEVPNGCTVIKINAQDLGNGD
IVTSFNDISPNVSKFHIDPTGQITVIGYIDFEBSKTEIIVGIDKQQLPLSGHC
RVIVEVEDNDNDVPLEFSLPIREDAPLGVIALISVSDKMGVNLVTCSTLSD
VPEKLSTFENYSLVLDALDRESAYELVVTADGSPSLMATAVSVEVADND
NAPAFQETVYVKKENPPGCHIFVSAAMDADQNALVSLVERVGERALSSYV
SVAESGKYVALQPLDHELELLOFVOTADACVPLGNSVTLQVFLDENAPALL
APRAGTGAVSELVPSVGVGHVAKVRVADDSGNMLSYELQPGGAGATIPRV
GLYTGELSTRALDETDPARHLVLYKHGEPAULATATVLSYVESCQAPASSRA
LVGAVGDAALVDVNYLITATCAVSSLLVLLLTALKCSALPTGCAKCPKPLV
CSSAVGMSYQQQRQVSCGEGPPTDLMAFSPSLPSDRDQDTQTTEFSAKPRQ
PNPDMRYASLRAGMHSVHLEAGILRAGGPGDQMPVSSATPEPEGEVSPVQ
AGVNSNSWTFYFGNPKOSGCELPKFTIPGSPALISIROEPTNSQIDKSDFTFG
GKEETKKKKKKKGNKQEKKEKGNSTTDSQ"

BASE COUNT 1360 a 1229 c 1304 g 1358 t
ORIGIN

Query Match 60.0%; Score 15; DB 89; Length 5251;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caagtcacagaatta 25
|||||

Db 4148 CAAGTCACAGAATTA 4134

RESULT 34
AF152306 5254 bp mRNA PRI 22-JUL-1999
LOCUS AF152306/c
DEFINITION Homo sapiens protocadherin alpha 10 (PCDH-alpha10) mRNA, complete
cDS
VERSION AF152306
KEYWORDS AF152306.1 GI:5456897
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS Wu,Q. and Maniatis,T.
TITLE A striking organization of a large family of human neural
cadherin-like cell adhesion genes
JOURNAL Cell 97 (6), 779-790 (1999)
MEDLINE 99308636
PUBMED 10380929
REFERENCE 2 (bases 1 to 5254)
AUTHORS Wu,Q. and Maniatis,T.
TITLE Human protocadherin genes
JOURNAL unpublished

REFERENCE 3 (bases 1 to 5254)
AUTHORS Wu,Q. and Maniatis,T.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-1999) Department of Molecular and Cellular
Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
02138, USA

FEATURES

source

Location/Qualifiers
1..5254

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="5"

/map="5q31"

/tissue-type="brain"

1..5254

/gene="PCDH-alpha10"

1..2847

/gene="PCDH-alpha10"

/function="cell-cell adhesion"

/product="protocadherin alpha 10"

/protein_id="AA043700.1"

/db_xref="GI:5456898"

gene
CDS

1..2847

/gene="PCDH-alpha10"

/function="cell-cell adhesion"

/product="protocadherin alpha 10"

/protein_id="AA043700.1"

/db_xref="GI:5456898"

/translation="WVSRSCGLVQVCLLSLLLAWEVSGQULHYVSEAKHCTE
GRIADQGLLELAEVLPRLFRVASKRGGLLEVNQNGILFVNSRIDRELCRSAC
SHELVYDPLQVHFVDEVDINDNPVFPATOKLISASRPDLSRFPLEGASDA
DIGNALTYRLSPNEYFLEKPPDEIKKGLILKSLDREAEIFLVATDGG
KPELCTGQVLLTVLDANNAFDPDTIKVRLLEVPNGCTVIKINAQDLGNGD
IVTSFNDISPNVSKFHIDPTGQITVIGYIDFEBSKTEIIVGIDKQQLPLSGHC
RVIVEVEDNDNDVPLEFSLPIREDAPLGVIALISVSDKMGVNLVTCSTLSD
VPEKLSTFENYSLVLDALDRESAYELVVTADGSPSLMATAVSVEVADND
NAPAFQETVYVKKENPPGCHIFVSAAMDADQNALVSLVERVGERALSSYV
SVAESGKYVALQPLDHELELLOFVOTADACVPLGNSVTLQVFLDENAPALLA
APRAGTGAVSELVPSVGVGHVAKVRVADDSGNMLSYELQPGGAGATIPRV
GLYTGELSTRALDETDPARHLVLYKHGEPAULATATVLSYVESCQAPASSRA
LVGAVGDAALVDVNYLITATCAVSSLLVLLLTALKCSALPTGCAKCPKPLV
CSSAVGMSYQQQRQVSCGEGPPTDLMAFSPSLPSDRDQDTQTTEFSAKPRQ
PNPDMRYASLRAGMHSVHLEAGILRAGGPGDQMPVSSATPEPEGEVSPVQ
AGVNSNSWTFYFGNPKOSGCELPKFTIPGSPALISIROEPTNSQIDKSDFTFG
GKEETKKKKKKKGNKQEKKEKGNSTTDSQ"

BASE COUNT 1358 a 1210 c 1317 g 1369 t
ORIGIN

Query Match 60.0%; Score 15; DB 89; Length 5254;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caagtcacagaatta 25
|||||

Db 4151 CAAGTCACAGAATTA 4137

RESULT 35
AF152310 5254 bp mRNA PRI 22-JUL-1999
LOCUS AF152310/c
DEFINITION Homo sapiens protocadherin alpha 2 (PCDH-alpha2) mRNA, complete
cDS
VERSION AF152310
KEYWORDS AF152310.1 GI:5456905
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS Wu,Q. and Maniatis,T.
TITLE A striking organization of a large family of human neural
cadherin-like cell adhesion genes
JOURNAL Cell 97 (6), 779-790 (1999)
MEDLINE 99308636
PUBMED 10380929
REFERENCE 2 (bases 1 to 5254)
AUTHORS Wu,Q. and Maniatis,T.

TITLE Human protocadherin genes
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 5254)
 AUTHORS Wu, Q. and Maniatis, T.
 TITLE Direct Submission
 JOURNAL Submitted (17-MAY-1999) Department of Molecular and Cellular
 Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
 02138, USA

FEATURES
 source
 Location/Qualifiers
 1..5254
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /map="5q31"
 /tissue_type="brain"
 1..5254
 /gene="PCDH-alpha2"
 1..2847
 /gene="PCDH-alpha2"
 /function="cell-cell adhesion"
 /codon_start=1
 /product="protocadherin alpha 2"
 /protein_id="AAD43704.1"
 /db_xref="GI:5456906"
 /translation="MASSIRGCGAMTRLLSLLLAMEVSGQLRYSPEAKHGTF
 VGRADOLGELIELVRLFRVASKRHGDLLENLQGLIFVNSRIDREPLCGSAC
 SIHEVATVDRPLOVFEHVEVKDINDPTEFPMIVKIRPEESRLDSRPLEGASDA
 DIGVNALSLKSLSSFEFLDIQANDESLSLVLGKSLDRETEAENVLLIATGCG
 KPELTGVOILLIVLDVNDNEPTFAQSVYKVLIENTANGTLVKNASADGPNSE
 IVLSGDSVSTIOTKFTIDPISGEIRTKGLDYEEAKSYEIOVATDKGTPMSGHC
 KISLKTVIDINDPEVSTISLPISENASLITVADIALITYSDRDSGNTGSLPH
 VPEKLVTEKNYSIVLDSALDREVSARELVYVARGGSPSLMATTSTVSEVADND
 NAFAPQEPYTVKKNPPCHITFTYSANDAOENALVSYLVERGERALSSYV
 SVARESKYALDPLDHEVELLQFOVSADACVPLGSNVLQVFLDENAPRL
 APRAGTAGAVSELPEWSVGAHVAKVADADSGYNMISYELQGTGSARIPRRV
 GLYTGELISTRALDEADSPRHRLIVLKHGEPALJATATVLSVESGAPKASSRA
 WYGAAGSEALIVVNYLLIACVSSILVLTALYLCVPTPEGARAPKPKLV
 CSSAVGMSYSQORORVCSGDEPKPTDLMAPSPSLSGDPSAEKQLSSFEVGVK
 QPNPDWYSASLRAGMSSVHLEAGILRAGPGPOOPTVSATPEPAGVSPV
 GAGYNSMSTFKYGPGRKSGGELPDKRTITGSPALLISIREPTNSQIDSKDFTTF
 GKKEETKKKKKKKNTOKKKGSTNDNSO"

BASE COUNT 1375 a 1234 c 1300 g 1345 t

ORIGIN

Query Match 60.0%; Score 15; DB 89; Length 5254;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caagtcacagaatta 25
 |||
 Db 4151 CAAGTCACAGAATTA 4137

RESULT 36
 AF152307 5257 bp mRNA PRI 22-JUL-1999
 LOCUS AF152307 Homo sapiens protocadherin alpha 11 (PCDH-alpha11) mRNA, complete
 DEFINITION cds.
 ACCESSION AF152307
 VERSION AF152307.1 GI:5456899
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 5257)
 AUTHORS Wu, Q. and Maniatis, T.
 TITLE A striking organization of a large family of human neural
 cadherin-like cell adhesion genes
 JOURNAL Cell 97 (6), 779-790 (1999)
 MEDLINE 99308636
 PUBMED 10380929

REFERENCE 2 (bases 1 to 5257)
 AUTHORS Wu, Q. and Maniatis, T.
 TITLE Human protocadherin genes
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 5257)
 AUTHORS Wu, Q. and Maniatis, T.
 TITLE Direct Submission
 JOURNAL Submitted (17-MAY-1999) Department of Molecular and Cellular
 Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
 02138, USA

FEATURES
 source
 Location/Qualifiers
 1..5257
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /map="5q31"
 /tissue_type="brain"
 1..5257
 /gene="PCDH-alpha11"
 1..2850
 /gene="PCDH-alpha11"
 /function="cell-cell adhesion"
 /codon_start=1
 /product="protocadherin alpha 11"
 /protein_id="AAD43701.1"
 /db_xref="GI:5456900"
 /translation="MEGRORGLGTPRLQMLLLEFEWEGSGQLHYSPREAHGTF
 VRIADOLGELIELVRLFRVASKRHGDLLENLQGLIFVNSRIDREPLCGSAC
 SIHEVATVDRPLOVFEHVEVKDINDPTEFPMIVKIRPEESRLDSRPLEGASDA
 DIGVNALSLKSLSSFEFLDIQANDESLSLVLGKSLDRETEAENVLLIATGCG
 KPELTGVOILLIVLDVNDNEPTFAQSVYKVLIENTANGTLVKNASADGPNSE
 IVLSGDSVSTIOTKFTIDPISGEIRTKGLDYEEAKSYEIOVATDKGTPMSGHC
 KISLKTVIDINDPEVSTISLPISENASLITVADIALITYSDRDSGNTGSLPH
 VPEKLVTEKNYSIVLDSALDREVSARELVYVARGGSPSLMATTSTVSEVADND
 NAFAPQEPYTVKKNPPCHITFTYSANDAOENALVSYLVERGERALSSYV
 SVARESKYALDPLDHEVELLQFOVSADACVPLGSNVLQVFLDENAPRL
 APRAGTAGAVSELPEWSVGAHVAKVADADSGYNMISYELQGTGSARIPRRV
 GLYTGELISTRALDEADSPRHRLIVLKHGEPALJATATVLSVESGAPKASSRA
 WYGAAGSEALIVVNYLLIACVSSILVLTALYLCVPTPEGARAPKPKLV
 CSSAVGMSYSQORORVCSGDEPKPTDLMAPSPSLSGDPSAEKQLSSFEVGVK
 QPNPDWYSASLRAGMSSVHLEAGILRAGPGPOOPTVSATPEPAGVSPV
 GAGYNSMSTFKYGPGRKSGGELPDKRTITGSPALLISIREPTNSQIDSKDFTTF
 GKKEETKKKKKKKNTOKKKGSTNDNSO"

BASE COUNT 1393 a 1224 c 1317 g 1323 t

ORIGIN

Query Match 60.0%; Score 15; DB 89; Length 5257;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caagtcacagaatta 25
 |||
 Db 4154 CAAGTCACAGAATTA 4140

RESULT 37
 AF152305 5260 bp mRNA PRI 22-JUL-1999
 LOCUS AF152305 Homo sapiens protocadherin alpha 1 (PCDH-alpha1) mRNA, complete
 DEFINITION cds.
 ACCESSION AF152305
 VERSION AF152305.1 GI:5456895
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 5260)
 AUTHORS Wu, Q. and Maniatis, T.
 TITLE A striking organization of a large family of human neural
 cadherin-like cell adhesion genes
 JOURNAL Cell 97 (6), 779-790 (1999)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 5260)
AUTHORS Wu, Q. and Maniatis, T.
TITLE A striking organization of a large family of human neural
cadherin-like cell adhesion genes

JOURNAL Cell 97 (6), 779-790 (1999)

MEDLINE 99308636
PUBMED 10380929

REFERENCE 2 (bases 1 to 5260)
AUTHORS Wu, Q. and Maniatis, T.
TITLE Human protocadherin genes
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 5260)
AUTHORS Wu, Q. and Maniatis, T.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-1999) Department of Molecular and Cellular
Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
02138, USA

FEATURES
source
1..5260
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q31"
/tissue_type="brain"
1..5260
/gene="PCDH-alpha8"
1..2853
/gene="PCDH-alpha8"
/function="cell-cell adhesion"
/product="protocadherin alpha 8"
/protein_id="A043710.1"
/db_xref="GI:5456920"
/translation="MDYHWGELGSLNLLLLLLLAAMVVGSGQLHYSPPEAKHGT
VGRADQGLLELAEVPRLEPRVASKRHRLDLEVSQNGILFVNSRLDRELCGRSAC
SIHLEIVDRPLOVHFVDEVDVNDNPNPVPRYKOKLFEVSERMPDSFPLEGSAD
DYGANVLTLYRLSSHDYFMDYNSKNDENKIVEYIKRSIDEDAPAHHLPTATDGG
KPELCTGTVOLLITVLDNNDNAVPDRITLYKPEVNSIGTLVHPNADSDDEGLND
ITSSGSSVSDISKPEMDPLSGAITYGIMDEESRAIKTPEANDKGPPLAGC
TLLEVVDVNDNAQOLITKITLSYVEKDAQLGVTALISYDLDADNGVOTSLTPH
VPFAFOSYKNYSLVLDRLADRESVAEELVVAROGGSPSLMATAVSLEYADVND
NAPFAOSEYVFEVKNPPGCHITFTVSARADAOENALVSYSILVERRLERSISYV
SVHASEGVAYALOPDLHELELLOFVSARADAGVPLGSNVTLOFVLDENDNAPLL
TPRRRGTVGAENVLRSVGVGVVAVRADADSGYNAMLSYELQFETAASIPPRV
GLTYGELSTTRVLDDEADSPRIHLVVKDHEPALTAIVLSVESQAKASSRO
SAGVLSPREALVDVNYLLIAICAVSSLLVLTALYLRCSALPTEGCGRAGKPTLV
CSSAVGSMYSQOORORVCSGEGPKTDLMARSPCLPDLGSDVGEEDLDVNDGLK
CPGPNPDMRYASASLRAGMHSVHLEAGILRAGPGGPDQOMPTVSSATPEEPAGVSP
PVGAGVNSMTFTFKYGPNGPKOSGPEGLPKFIIPGSPALISIRQEPITNSQIDKSDFI
TFGKKEETKTKKKKKGKNTQEKKEKGNSTTNSDQ"

BASE COUNT 1357 a 1236 c 1310 g 1357 t

ORIGIN

Query Match 60.0%; Score 15; DB 89; Length 5260;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaatta 25
|||||
Db 4157 CAAGTCACAGAAATTA 4143

RESULT 42
AF152317/c 5260 bp mRNA PRI 22-JUL-1999
LOCUS Homo sapiens protocadherin alpha 9 (PCDH-alpha9) mRNA, complete
DEFINITION
ACCESSION AF152317
VERSION AF152317.1 GI:5456921

KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 5260)
AUTHORS Wu, Q. and Maniatis, T.
TITLE A striking organization of a large family of human neural
cadherin-like cell adhesion genes

JOURNAL Cell 97 (6), 779-790 (1999)

MEDLINE 99308636
PUBMED 10380929

REFERENCE 2 (bases 1 to 5260)
AUTHORS Wu, Q. and Maniatis, T.
TITLE Human protocadherin genes
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 5260)
AUTHORS Wu, Q. and Maniatis, T.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-1999) Department of Molecular and Cellular
Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
02138, USA

FEATURES
source
1..5260
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q31"
/tissue_type="brain"
1..5260
/gene="PCDH-alpha9"
1..2853
/gene="PCDH-alpha9"
/function="cell-cell adhesion"
/product="protocadherin alpha 9"
/protein_id="A043711.1"
/db_xref="GI:5456922"
/translation="MLYSSRQDPGQPLLSLLILAMVVGSGQLHYSPPEAEHGT
VGRADQGLLELAEVPRLEPRVASKRHRLDLEVSQNGILFVNSRLDRELCGRSAC
SIHLEIVDRPLOVHFVDEVDVNDNPNPVPRYKOKLFEVSERMPDSFPLEGSAD
DYGANVLTLYRLSSHDYFMDYNSKNDENKIVEYIKRSIDEDAPAHHLPTATDGG
KPELCTGTVOLLITVLDNNDNAVPDRITLYKPEVNSIGTLVHPNADSDDEGLND
ITSSGSSVSDISKPEMDPLSGAITYGIMDEESRAIKTPEANDKGPPLAGC
TLLEVVDVNDNAQOLITKITLSYVEKDAQLGVTALISYDLDADNGVOTSLTPH
VPFAFOSYKNYSLVLDRLADRESVAEELVVAROGGSPSLMATAVSLEYADVND
NAPFAOSEYVFEVKNPPGCHITFTVSARADAOENALVSYSILVERRLERSISYV
SVHASEGVAYALOPDLHELELLOFVSARADAGVPLGSNVTLOFVLDENDNAPLL
TPRRRGTVGAENVLRSVGVGVVAVRADADSGYNAMLSYELQFETAASIPPRV
GLTYGELSTTRVLDDEADSPRIHLVVKDHEPALTAIVLSVESQAKASSRA
SAGVLSPREALVDVNYLLIAICAVSSLLVLTALYLRCSALPTEGCGRAGKPTLV
CSSAVGSMYSQOORORVCSGEGPKTDLMARSPCLPDLGSDVGEEDLDVNDGLK
CPGPNPDMRYASASLRAGMHSVHLEAGILRAGPGGPDQOMPTVSSATPEEPAGVSP
PVGAGVNSMTFTFKYGPNGPKOSGPEGLPKFIIPGSPALISIRQEPITNSQIDKSDFI
TFGKKEETKTKKKKKGKNTQEKKEKGNSTTNSDQ"

BASE COUNT 1354 a 1257 c 1320 g 1329 t

ORIGIN

Query Match 60.0%; Score 15; DB 89; Length 5260;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaatta 25
|||||
Db 4157 CAAGTCACAGAAATTA 4143

RESULT 43
AF152303/c 5299 bp mRNA PRI 22-JUL-1999
LOCUS Homo sapiens protocadherin alpha C1 (PCDH-alpha-C1) mRNA, complete
DEFINITION
ACCESSION AF152303
VERSION AF152303.1 GI:5456921

[illegible]

DEFINITION	cds.	Homo sapiens protocadherin alpha C2 (PCDH-alpha-C2) mRNA, complete cds.
ACCESSION		AF152304
VERSION		AF152304.1
KEYWORDS		GI:5456893
SOURCE		human.
ORGANISM		Homo sapiens
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE		1 (bases 1 to 5431)
TITLE		Wu,Q. and Maniatis,T.
JOURNAL		A striking organization of a large family of human neural
MEDLINE		cadherin-like cell adhesion genes
PUBMED		Cell 97 (6), 779-790 (1999)
REFERENCE		2 (bases 1 to 5431)
AUTHORS		Wu,Q. and Maniatis,T.
TITLE		Human protocadherin genes
JOURNAL		Unpublished
REFERENCE		3 (bases 1 to 5431)
AUTHORS		Wu,Q. and Maniatis,T.
TITLE		Direct Submission
JOURNAL		Submitted (17-May-1999) Department of Molecular and Cellular
FEATURES		Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
source		02138, USA
Location/Qualifiers		
1..5431		
/organism="Homo sapiens"		
/db_xref="taxon:9606"		
/chromosome="5"		
/map="5q31"		
/tissue_type="brain"		
1..5431		
/gene="PCDH-alpha-C2"		
1..3024		
/gene="PCDH-alpha-C2"		
/function="cell-cell adhesion"		
/codon_start=1		
/product="protocadherin alpha C2"		
/protein_id="AADA3698.1"		
/db_xref="GI:5456894"		
/translation="MEQAGRPATEHERRLRRPWWLLPLLLLLLLLLLPGPAASQ		
RSVSEOPALGVNVARALGELRLGCGCLRIHNLGAPSPRYLEDLTSGLFVNA		
PRIDREALCEORPCLISLEYLAHNPVAVSAVEVELIDNDSPPRPNTOLGYSEE		
VARGARHIESADNPDCANSVOTYELSPSEHRELKPKLOENSVYLELVRKGDRE		
QALHLVLAVDGGITPARSGTQGISRYLDTNDNSPARDQSTRYVQLPESPEPTL		
VNLNSDDEGSNGELRYSYSTSDERQLSIDSTGEVARTIGDLTEASSIQIY		
VOATRGVPVPAHGCKYLVLDIVYDNAPPVVLTLDYSPVENATPNTVAALVND		
DSGPRKYSGLIEALTLPFLRLNGFNSYTLVSGPLDRERAVNIATVATDGIQL		
SIRTLKVELSIDNDNPSEPLEDSYSITYIOENNLPGVLLCTVOATPDEKENAEVY		
LERETIOGLPVTSYVSINSAGSLAAVNSFPYKEREFPYVEADOKSGPSSITYTA		
VYVDNDHAPHILLYPTNSGSLAFEMVPTPAAGLYLTKVLAOMDSQNMFLFYH		
AOTSLDLIKVELHTELRTTRKMGDESGTFNLYVVDNGSPISLSAVATVAVV		
RSKSLTPTORHAKVSESTLYLILALSTVSFTFLFLITLITIKCYRTTAGT		
CCGCGCGERSPAELYKOANNITDARIPIGLAVOPHFLEVKNQNSLTKTYCKACL		
AASSGDTFMFYNTGAQTGPSPGSAQAAVTDTSRLTQSGSQMAGNLITLKNEVSN		
ROPNDWRTSASLRAGHMSVYLEEAGILKAGGCGDDQMPYVSATPEPAGVSPYS		
VAGVNSWMTFKYCPGCKPQSGPELPEKFIIPGSPALISIRQEPNTSOLQSDFI		
PKKKETKTKKKKKKKGKNTQKKKKGGSTTNDSDQ"		

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:40:39 ; Search time 469.84 Seconds
(without alignments)
33.410 Million cell updates/sec

Title: US-09-396-196F-1

Perfect score: 25
1 gacatgtgcgaagtcacagaatta 25

Scoring table:
OLIGO-NUC
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 9

Total number of hits satisfying chosen parameters: 24134

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

N_Geneseq_0601:*

- 1: /SIDSI/gcgdata/geneseq/geneseqn/NA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseqn/NA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseqn/NA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseqn/NA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseqn/NA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseqn/NA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseqn/NA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseqn/NA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseqn/NA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseqn/NA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseqn/NA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseqn/NA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseqn/NA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseqn/NA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseqn/NA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseqn/NA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseqn/NA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseqn/NA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	1041	20	AA01303
2	25	100.0	1084	10	AA01329
3	25	100.0	1121	7	AA060496
4	25	100.0	5872	15	AA062386
5	17	68.0	17	14	AA056307
6	15	60.0	561	22	AA094235
7	15	60.0	1771	20	AA230602
8	15	60.0	3520	22	AA093909
9	15	60.0	3520	22	AA029361
10	15	60.0	6567	11	AA003324
11	14	56.0	1050	9	AA080927

C 12	14	56.0	1677	17	AA013932
C 13	14	56.0	1787	17	AA013928
C 14	14	56.0	1890	20	AA052940
C 15	14	56.0	2233	19	AA010120
C 16	14	56.0	2585	22	AA077541
C 17	14	56.0	3083	12	AA011851
C 18	14	56.0	5773	21	AA026014
C 19	14	56.0	32351	21	AA021307
C 20	14	56.0	32351	21	AA035185
C 21	14	56.0	40238	21	AA021311
C 22	14	56.0	40298	21	AA035189
C 23	14	56.0	1038602	20	AA0201425
C 24	13	52.0	87	16	AA022576
C 25	13	52.0	167	16	AA020233
C 26	13	52.0	187	21	AA026213
C 27	13	52.0	363	21	AA050159
C 28	13	52.0	378	13	AA025667
C 29	13	52.0	408	21	AA074604
C 30	13	52.0	415	21	AA056783
C 31	13	52.0	476	21	AA009392
C 32	13	52.0	559	22	AA093531
C 33	13	52.0	637	21	AA095830
C 34	13	52.0	713	15	AA081957
C 35	13	52.0	777	18	AA079355
C 36	13	52.0	778	20	AA032456
C 37	13	52.0	902	21	AA056756
C 38	13	52.0	930	20	AA015335
C 39	13	52.0	1002	21	AA050164
C 40	13	52.0	1010	21	AA064778
C 41	13	52.0	1010	21	AA027678
C 42	13	52.0	1278	21	AA036039
C 43	13	52.0	1317	21	AA094840
C 44	13	52.0	1359	21	AA072281
C 45	13	52.0	1404	21	AA059716

ALIGNMENTS

RESULT 1	
AA01303	AA01303 standard; DNA; 1041 BP.
XX	XX
XX	AA01303;
XX	XX
DT	12-APR-1999 (first entry)
XX	XX
DE	E. coli biotin synthetase (Biot) coding sequence.
XX	XX
KW	DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KM	biotin synthetase; biotin production; vitamin H; Biot; ss.
XX	XX
OS	Escherichia coli.
XX	XX
PN	US5869719-A.
XX	XX
PD	09-FEB-1999.
XX	XX
PF	30-APR-1997; 97US-0846338.
XX	XX
PR	30-APR-1997; 97US-0846338.
PR	08-MAR-1995; 95US-0401068.
XX	XX
PA	(NOVS) NOVARTIS FINANCE CORP.
XX	XX
PI	Patton DA;
XX	XX
DR	WPI; 1999-152902/13.
XX	XX
DR	P-PSDB; AA073906.
XX	XX
PT	Transgenic plants with high biotin levels - transformed with DNA
PT	encoding di:amino-pelargonic acid amino-transferase or biotin
PT	synthase

Retinoid X recepto
Retinoid X recepto
Human prostate tum
Human retinoid rec
Caenorhabditis bri
Glutamate receptor
Human ORF XRF1569
Human low adenosin
Human adenosine re
Human low adenosin
Human adenosine re
Complete genome se
Human gene signatu
Human gene signatu
Human secreted pro
Antibody 5H7 heavy
Sequence of the an
Murine 5B3 antibod
Euclalyptus grandis
Human secreted pro
CDNA encoding SRT
Bovine metalloprot
C22(487-713). Ca
Staphylococcus aur
Partial nucleotide
Human transmembran
CDNA encoding a le
DNA encoding 5H7 s
Strawberry alcohol
Arabidopsis thaliana
Drosophila melanog
Drosophila odorant
DNA encoding a mut

XX Example 2; Column 37-40; 34pp; English.

PS This sequence encodes the E. coli biotin synthetase (BiotB). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.

XX Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

QY Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. NO. 7.2e-06;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 gacattgctgcaagtcacagaatta 25
 21 gacattgctgcaagtcacagaatta 45

RESULT 2

ID AAN91329 standard; DNA; 1084 BP.

AC AAN91329;

DT 15-FEB-1990 (first entry)

DE E.coli B10 B gene.

KW E.coli; B10 B gene; biotin.

OS Escherichia coli.

XX Key Location/Qualifiers

FT CDS 24..1064

PN GB2216530-A.

PD 11-OCT-1989.

PF 17-MAR-1989; 89GB-0006210.

PR 22-MAR-1988; 88GB-0006804.

PR 17-MAR-1989; 89GB-0006210.

PA (UKAG-) UK MIN. AGRIC. FISHER.

PI Pearson BM, McKee RA;

DR WPI; 1989-295085/41. P-PSDB P91392

PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes
 derived from E.coli and capable of replication and expression in other
 microorganisms, esp. yeast.

PS Table 3; page 33-4; 52pp; English.

CC The gene can be used in a plasmid for expression of enzymes of the biotin
 CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
 CC are plasmids pMA91, pMA36c, pKY49 and pCK495, and plasmid pCK965 for
 CC Lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.

XX Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

QY Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. NO. 7.2e-06;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 gacattgctgcaagtcacagaatta 25
 44 gacattgctgcaagtcacagaatta 68

RESULT 3

ID AAN60496 standard; DNA; 1121 BP.

AC AAN60496;

DT 17-OCT-1991 (first entry)

DE Sequence encoding biotin synthesising enzyme.

KW Biotin synthetic enzyme; E.coli; deslbiobiotin; ds.

XX Key Location/Qualifiers

FT CDS 42..1082

PN JP61149091-A.

PD 07-JUL-1986.

PF 24-DEC-1984; 84JP-0272605.

PR 24-DEC-1984; 84JP-0272605.

PA (NIPS) NIPPON SODA KK.

DR WPI; 1986-216622/33.

DR P-PSDB; AAN60536.

PT Double stranded DNA encoding biotin synthesising enzyme -
 complises transformed mutant E.coli strain contg. cyclic doubled
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.

PS Disclosure; Page 534; 23pp; Japanese.

CC The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing deslbiobiotin.

XX Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

QY Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. NO. 7.2e-06;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 gacattgctgcaagtcacagaatta 25
 62 gacattgctgcaagtcacagaatta 86

RESULT 4

ID AAO62386 standard; DNA; 5872 BP.

AC AAO62386;

DT 16-NOV-1994 (first entry)

DE Biotin-biosynthesis genes contg. plasmid PB030A-15/9.

KW Biotin; expression; enterobacteria; vitamin H; synthesis;

KW plasmid; PB030A-15/9; bioB; bioC; bioD; bioA;

KW promoter; plac; biotin synthase; KAPA synthase;

KW 8-amino-7-oxononanoate synthase; pimeloyl-CoA; DFB synthase;


```

XX 31-AUG-1993.
PD
XX
XX 14-SEP-1992; 92JP-0244792.
PF
XX 14-SEP-1992; 92JP-0244792.
PR
XX (SHS ) SHISEIDO CO LTD.
PA (TAKE ) TAKEDA CHEM IND LTD.
XX
XX WPI; 1993-308323/39.
DR
XX
XX DNA sequence of biotin operon - has base sequence of E. coli
PT mutated by base pair(s) compared to wild type
XX
XX Example 1; Fig 8; 11pp; Japanese.
PS
XX
XX A novel DNA sequence comprises the E.coli biotin operon (BO) in which
CC the control region of BO or the region near the biob initiation
CC codon is mutated by at least one base pair compared to its
CC wild type. Two primers (AA056306-Q56307) are described in Example 1.
CC A microorganism belonging to Escherichia genus, transformed by
CC a recombinant plasmid carrying such DNA can be used for the prodn.
CC of biotin-active substances.
XX
XX Sequence 17 BP; 3 A; 4 C; 4 G; 6 T; 0 other;
SQ

Query Match 68.0%; Score 17; DB 14; Length 17;
Best Local Similarity 100.0%; Pred. No. 0.34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 gtcgcaagtcacagaat 23
    |||||||||||||||
Db 17 GTCCCAAGTCACAGAA 1

RESULT 6
AAF94235
ID AAF94235 standard; DNA; 561 BP.
XX
XX AAF94235;
AC
XX
XX 23-MAY-2001 (first entry)
DT
XX
XX Primer specific for DNA encoding secretory/membrane protein SEQ ID 669.
DE
XX
XX Human; secretory protein; membrane protein; vaccine; gene therapy;
KW rheumatoid arthritis; diabetes; PCR primer; ss.
XX
XX Synthetic.
OS
XX
XX EP1067182-A2.
PN
XX
XX 10-JAN-2001.
PD
XX
XX 07-JUL-2000; 2000EP-0114090.
PF
XX
XX 08-JUL-1999; 99JP-0194179.
PR 11-JAN-2000; 2000JP-0118775.
PR 02-MAY-2000; 2000JP-0183766.
XX
XX (HELI-) HELIX RES INST.
PA
XX
XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
PI WPI; 2001-093989/11.
DR
XX
XX Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT gene therapy or as candidate target molecules in drug development -
XX
XX Claim 5; SEQ ID 669; 609bp + CD ROM; English.
PS
XX

```

```

CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by
CC AAF88317 - AAF88419. Included in the invention are primers
CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC cDNA sequences of the invention. The invention also includes methods for
CC the production of antibodies directed against the proteins, and cDNA
CC sequences, which can be used in vaccines. The polynucleotide sequences
CC can be used in gene therapy. The polynucleotide sequences and the
CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with inappropriate secretory
CC protein/membrane protein expression. The nucleic acids and complementary
CC sequences may also be used as DNA probes in diagnostic assays
CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC presence of similar nucleic acid sequences in samples. They may also be
CC used to study the expression and function of secretory proteins/membrane
CC polypeptides and their role in metabolism. The polypeptides may be used
CC as antigens in the production of antibodies against them and in assays to
CC identify modulators (agonists and antagonists) of expression and
CC activity. The antibodies and antagonists may also be used as therapeutic
CC agents to down regulate expression and activity. The antibodies may also
CC be used as diagnostic agents for detecting the presence of the
CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC (ELISA)). Examples of diseases which may be treated include rheumatoid
CC arthritis and diabetes.
XX
XX Sequence 561 BP; 213 A; 83 C; 114 G; 149 T; 2 other;
SQ

Query Match 60.0%; Score 15; DB 22; Length 561;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 caagtcacagaatta 25
    |||||||||||||||
Db 339 caagtcacagaatta 353

RESULT 7
AAZ30602
ID AAZ30602 standard; DNA; 1771 BP.
XX
XX AAZ30602;
AC
XX
XX 18-JAN-2000 (first entry)
DT
XX
XX Mouse integrin alpha 4 coding sequence.
DE
XX
XX Human; integrin; antisense; oligonucleotide; inhibition; expression;
KW very late antigen; CD49d; CD29; cell surface; leucocyte; adhesion;
KW vascular endothelial cell; vascular endothelium; migration; inflammation;
KW atherosclerosis; allergy; asthma; rheumatoid arthritis; tumor;
KW metastasis; circulatory system; autoimmune disease; Grave's disease;
KW Hashimoto's thyroiditis; encephalomyelitis; multiple sclerosis; ds.
XX
XX Mus sp.
OS
XX
XX US5968826-A.
PN
XX
XX 19-OCT-1999.
PD
XX
XX 05-OCT-1998; 98US-0166203.
PF
XX
XX 05-OCT-1998; 98US-0166203.
PR
XX
XX 05-OCT-1998; 98US-0166203.
XX
XX (ISIS-) ISIS PHARM INC.
PA
XX
XX Bennett CF, Cowsett LM, Condon TP;
PI WPI; 1999-590416/50.
DR
XX
XX P-PDB; AA44154.
DR
XX
XX Antisense inhibition of integrin alpha4 expression useful for treating
PT inflammatory diseases such as atherosclerosis, allergies, asthma and

```

PT arthritis -
 XX
 PS Example 12; Column 71-74; 40pp; English.
 XX
 CC This sequence represents the coding region of the mouse integrin
 CC alpha4 gene. The invention relates to the generation of antisense
 CC oligonucleotides targeted to the integrin alpha4 gene which are used
 CC for inhibiting expression of the integrin alpha4 mRNA or protein.
 CC Integrin alpha4 is a component of Very Late Antigen (VLA)-4 (also
 CC called alpha4beta1 and CD49d/CD29). VLA-4 is expressed on the cell
 CC surfaces of leucocytes and vascular endothelial cells and mediates the
 CC adhesion of leucocytes to the vascular endothelium prior to migration
 CC into the surrounding tissues. This migration is an essential step in
 CC inflammation and hence VLA-4 (and consequently integrin alpha4) is a
 CC potential therapeutic target for treating inflammatory diseases and
 CC the damaging effects of excessive inflammation. These disorders include
 CC atherosclerosis, allergies, asthma, rheumatoid arthritis and tumor cell
 CC metastasis (VLA-4 is involved in migration of the tumor cells through
 CC the extracellular matrix into the circulatory system). VLA-4 is also
 CC involved in a number of autoimmune diseases such as Grave's disease,
 CC Hashimoto's thyroiditis, encephalomyelitis (EAF), multiple sclerosis,
 CC VLA-4 may also be involved in promoting adhesion (i.e. retention) of
 CC hemopoietic stem cells in bone-marrow and in allograft rejection.
 CC
 XX
 SQ Sequence 1771 BP; 371 A; 471 C; 519 G; 410 T; 0 other;

Query Match 60.0%; Score 15; DB 20; Length 1771;
 Best Local Similarity 100.0%; Pred. No. 5.1;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gcaagtcacagaatt 24
 |||
 Db 312 gcaagtcacagaatt 326

RESULT 8
 AAF93909/c
 ID AAF93909 standard; cDNA; 3520 BP.

XX AAF93909;

DT 23-MAY-2001 (first entry)

DE Human cDNA encoding a membrane or secretory protein clone PSEC0256.

XX Human; secretory protein; membrane protein; vaccine; gene therapy;

KW Rheumatoid arthritis; diabetes; ss.

XX Homo sapiens.

PN EP1067182-A2.

PD 10-JAN-2001.

PF 07-JUL-2000; 2000EP-0114090.

PR 08-JUL-1999; 99JP-0194179.

PR 11-JAN-2000; 2000JP-0118775.

PR 02-MAY-2000; 2000JP-0183766.

XX (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;

DR WPI; 2001-093989/11.

XX P-PSDB; AAB88482.

PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
 PT gene therapy or as candidate target molecules in drug development -
 XX
 PS Claim 1; SEQ ID 331; 609pp + CD ROM; English.
 XX

CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
 CC which encode human secretory or membrane proteins represented by
 CC AAB88317 - AAB88419. Included in the invention are primers
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
 CC cDNA sequences of the invention. The invention also includes methods for
 CC the production of antibodies directed against the proteins, and cDNA
 CC sequences, which can be used in vaccines. The polynucleotide sequences
 CC can be used in gene therapy. The polynucleotide sequences and the
 CC proteins they encode may be used in the prevention, treatment and
 CC diagnosis of diseases associated with inappropriate secretory
 CC protein/membrane protein expression. The nucleic acids and complementary
 CC sequences may also be used as DNA probes in diagnostic assays
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
 CC presence of similar nucleic acid sequences in samples. They may also be
 CC used to study the expression and function of secretory proteins/membrane
 CC polypeptides and their role in metabolism. The polypeptides may be used
 CC as antigens in the production of antibodies against them and in assays to
 CC identify modulators (agonists and antagonists) of expression and
 CC activity. The antibodies and antagonists may also be used as therapeutic.
 CC agents to down regulate expression and activity. The antibodies may also
 CC be used as diagnostic agents for detecting the presence of the
 CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay
 CC (ELISA)). Examples of diseases which may be treated include rheumatoid
 CC arthritis and diabetes.
 CC
 XX
 SQ Sequence 3520 BP; 871 A; 896 C; 901 G; 852 T; 0 other;

Query Match 60.0%; Score 15; DB 22; Length 3520;
 Best Local Similarity 100.0%; Pred. No. 5.1;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caagtcacagaatta 25
 |||
 Db 3182 CAAGTCACAGAAATTA 3168

RESULT 9
 AAF29361/c
 ID AAF29361 standard; DNA; 3520 BP.

XX AAF29361;

DT 20-APR-2001 (first entry)

DE Amyloid-beta protein agglutination regulating factor DNA SEQ ID 9.

XX Human; amyloid-beta protein; agglutination regulatory factor;

KW Alzheimer's disease; ds.

XX Homo sapiens.

PN WO200104299-A1.

PD 18-JAN-2001.

PF 06-JUL-2000; 2000WO-JP04515.

PR 08-JUL-1999; 99JP-0194179.

PR 18-OCT-1999; 99US-0159586.

XX (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Kawai Y, Yamazaki M, Satoh S;

DR WPI; 2001-138347/14.

XX P-PSDB; AAB49771.

PT Polynucleotide encoding Amyloid-beta protein agglutination-controlling
 PT factor, useful for inhibiting or promoting agglutination or
 PT sedimentation of amyloid-beta protein and in diagnosis and screening
 PT drugs for Alzheimer's disease -
 XX

XX Claim 1; Page 58-63; 72pp; Japanese.
 PS
 CC This invention relates to polynucleotides AAF29357 - AAF29361 which
 CC encode proteins AAB49767 - AAB49771. The proteins inhibit or promote the
 CC agglutination of amyloid beta protein. The proteins and polynucleotide
 CC sequences are useful in the diagnosis of Alzheimer's disease. They are
 CC also useful for screening drugs which are useful for treating Alzheimer's
 CC disease.
 XX
 SQ Sequence 3520 BP; 871 A; 896 C; 901 G; 852 T; 0 other;
 Query Match 60.0%; Score 15; DB 22; Length 3520;
 Best Local Similarity 100.0%; Pred. No. 5.1;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 caagtcacagaata 25
 |||||
 Db 3182 CAAGTCACAGAAATTA 3168
 RESULT 10
 ID AAO03324/c
 XX AAO03324 standard; DNA; 6567 BP.
 AC
 XX AAO03324;
 DT 02-AUG-1990 (first entry)
 XX
 DE Elmeria tenella genomic DNA encoding antigen GX5401FL.
 XX
 KM Elmeria tenella; antigen GX5401FL; antigen GX5401; avian coccidiosis.
 XX
 OS Elmeria tenella.
 XX
 FH Key Location/Qualifiers
 FT 1..6567
 FT CDS /tag= a
 XX
 PN WO9000403-A.
 XX
 PD 25-JAN-1990.
 XX
 PF 05-JUL-1989; 89WO-US02918.
 XX
 PR 05-JUL-1988; 88US-0215162.
 XX
 PA (GENE-) GENEX CORP.
 PI
 PT Anderson DM, McCandliss RJ, Strausberg SL, Strausberg RL;
 DR WPI: 1990-051586/07.
 DR P-PSDB; AAR05222.
 XX
 PT Cloned gene or fragment encoding antigenic protein -
 PT which binds with antibodies against avian coccidia, and
 PT transformed cells used in vaccine
 PS
 PS Claim 10; Page 93 and Fig 14; 134pp; English.
 CC
 CC An E. tenella genomic library was screened with radioactively labeled
 CC cDNA encoding the GX5401 antigen to identify clones encoding extensions
 CC of the GX5401 coding sequence. Several plaques reactive with the cDNA
 CC were identified. DNA from clone number 533 was sequenced. This analysis
 CC showed that the coding sequence in this clone overlaps with the GX5401
 CC coding sequence and extends that sequence toward the N-terminal coding
 CC region. Analysis of the new sequence together with the GX5401 antigen
 CC coding sequence reveals an open reading frame encoding an Elmeria
 CC protein of about 250,000 daltons. The protein carries several repeated
 CC peptide sequences and is rich in cysteine residues. The open reading
 CC frame encodes a potential signal sequence for protein secretion. The
 CC coding sequence for this antigenic protein is given in AAN93324. Also new

CC are an expression vector contg. the cloned gene, host cells transformed
 CC with the vector, and AP encoded by the cloned gene. The transformed cells
 CC are used in a vaccine to immunise birds against avian coccidiosis. By
 CC labelling the peptides, they can be used as a type-specific probe. The
 CC AP may also be used in an assay to detect Ab against the coccidia. The
 CC Abs are used to identify transformed cells contg. the DNA.
 XX
 SQ Sequence 6567 BP; 1490 A; 1741 C; 2189 G; 1147 T; 0 other;
 Query Match 60.0%; Score 15; DB 11; Length 6567;
 Best Local Similarity 100.0%; Pred. No. 5.1;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 gacattgctcgaagt 15
 |||||
 Db 3486 GACATTGTCGCAAGT 3472
 RESULT 11
 ID AAN80927/c
 XX AAN80927 standard; DNA; 1050 BP.
 AC
 XX AAN80927;
 DT 15-OCT-1990 (first entry)
 XX
 DE Lividans Exported Protein gene.
 XX
 KM Longisporus trypsin inhibitor; Lividans exported protein; LTI; LEP-10;
 XX
 KM protease inhibitor; probe; ss.
 XX
 OS Streptomyces lividans strain 1326.
 XX
 FH Key Location/Qualifiers
 FT sig.peptide 394..498
 FT mat.peptide 499..825
 FT /tag= b
 FT /product=LEP-10
 XX
 PN WO8801278-A.
 XX
 PD 25-FEB-1988.
 XX
 PF 17-AUG-1987; 87WO-US02009.
 XX
 PR 18-AUG-1986; 86US-0897245.
 XX
 PA (SMIR) SMITHKLINE BECKMAN CORP.
 PA (BERK/) TR BERKA.
 PI
 PT Berka TR, Fornwald JA, Gorniak JG, Rosenberg M;
 DR WPI: 1988-06396/09.
 DR P-PSDB; AAP80944.
 XX
 PT Novel protease inhibitors LEP-10 and LTI - produced by Streptomyces spp.
 PT disclosure; ; pp; English.
 PS
 PS The sequence can be cloned and used to produce recombinant LEP-10.
 CC The DNA can be used as a probe for identifying further coding
 CC sequences for protease inhibitors; the protein is useful as an
 CC assay reagent.
 CC See also AAN80926.
 XX
 SQ Sequence 1050 BP; 149 A; 398 C; 366 G; 137 T; 0 other;
 Query Match 56.0%; Score 14; DB 9; Length 1050;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aagtcacagaatta 25
 |||||||
 DB 251 AAGTCACAGAAATTA 238

RESULT 12

AAT31932/C
 ID AAT31932 standard; DNA; 1677 BP.

AC AAT31932;

XX 27-SEP-1996 (first entry)

DE Retinoid X receptor interacting protein RIP14-2 DNA.

XX Retinoid X receptor interacting protein; RXR; RIP; RIP14-2; ss.

XX Mus sp.

XX Key Location/Qualifiers

FT CDS 1..1356

FT polyA_signal 1672-1677

FT /*tag= b

PN WO9621677-A1.

PD 18-JUL-1996.

PF 08-DEC-1995; 95WO-US16311.

XX 13-JAN-1995; 95US-0372652.

XX (GEHO) GEN HOSPITAL CORP.

PI Choi H, Moore D, Seol W;

XX WPI; 1996-342241/34.

DR P-PSDB; AAR99735.

XX Retinoid X receptor (RXR) interacting protein (RIP) - useful to

PT modulate or mediate RXR function, anti-RIP antibodies can be used to

XX determine RIP subcellular distribution patterns

PS Claim 8; Page 59-60; 90pp; English.

XX A cDNA clone (AAT31932) codes for retinoid X receptor interacting

CC protein RIP14-2 (AAR99735), a previously undescribed orphan member

CC of the nuclear receptor superfamily which may be involved in the

CC complex retinoid response. It was isolated from a mouse liver

CC cDNA library using an RIP14 probe. Different RIP14-2 clones had

CC different 5' sequences (see also AAT31935-37). Another isoform,

CC RIP14-1, was also detected (see also AAT31928). The cDNA can be

CC used for the recombinant prodn. of RIP14-2 in transformed host

CC cells.

XX Sequence 1677 BP; 486 A; 385 C; 408 G; 398 T; 0 other;

XX Query Match 56.0%; Score 14; DB 17; Length 1677;

XX Best Local Similarity 100.0%; Pred. No. 20;

XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX OY 5 ttgtgcgaagtcac 18

DB 630 TTGTGCGCAAGTCAC 617

RESULT 13

AAT31928/C

ID AAT31928 standard; DNA; 1787 BP.

XX

AC AAT31928;

XX 27-SEP-1996 (first entry)

DE Retinoid X receptor interacting protein RIP14-1 DNA.

XX Retinoid X receptor interacting protein; RXR; RIP; RIP14-1; ss.

XX Mus sp.

XX Key Location/Qualifiers

FT CDS 12..1466

FT polyA_signal 1782..1787

FT /*tag= b

PN WO9621677-A1.

PD 18-JUL-1996.

PF 08-DEC-1995; 95WO-US16311.

XX 13-JAN-1995; 95US-0372652.

XX (GEHO) GEN HOSPITAL CORP.

PI Choi H, Moore D, Seol W;

XX WPI; 1996-342241/34.

DR P-PSDB; AAR99735.

XX Retinoid X receptor (RXR) interacting protein (RIP) - useful to

PT modulate or mediate RXR function, anti-RIP antibodies can be used to

XX determine RIP subcellular distribution patterns

PS Claim 8; Page 53-54; 90pp; English.

XX A full-length cDNA clone (AAT31928) codes for retinoid X receptor

CC interacting protein RIP14-1 (AAR99735), a previously undescribed

CC orphan member of the nuclear receptor superfamily which may be

CC involved in the complex retinoid response. It was isolated from

CC a mouse liver cDNA library using an RIP14 probe. Another isoform,

CC RIP14-2, was also detected (see also AAT31932). The cDNA can be used

CC for the recombinant prodn. of RIP14-1 in transformed host cells.

XX Sequence 1787 BP; 515 A; 408 C; 442 G; 422 T; 0 other;

XX Query Match 56.0%; Score 14; DB 17; Length 1787;

XX Best Local Similarity 100.0%; Pred. No. 20;

XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX OY 5 ttgtgcgaagtcac 18

DB 740 TTGTGCGCAAGTCAC 727

RESULT 14

AA252940/C

ID AA252940 standard; cDNA; 1890 BP.

XX AA252940;

XX 14-MAR-2000 (first entry)

DE Human prostate tumor cDNA library derived EST fragment #83.

XX Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;

XX treatment; ds.

XX Homo sapiens.

OS

XX DE19820190-A1.

PN

XX 04-NOV-1999.
 PD XX
 XX 28-APR-1998; 98DE-1020190.
 PF XX
 XX 28-APR-1998; 98DE-1020190.
 PR XX
 XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
 PA XX
 XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
 PI XX
 XX WPI; 1999-621386/54.
 DR XX
 XX P-PSDB; AAV74061, AAV74062, AAV74063.
 XX
 PT New human nucleic acid sequences from pancreatic tumors, and related
 PT proteins -
 XX
 PS Claim 2; Page 250; 502pp; German.
 XX
 CC This invention describes novel polypeptides and their encoding nucleic
 CC acids derived from human pancreatic tumor tissue which have cytostatic
 CC activity. The sequences are also useful in producing pharmaceutical
 CC compositions for treatment of pancreatic tumors. AA252858-253014
 CC represent expressed sequence tag (EST) fragments derived from a human
 CC pancreatic tumor cDNA library and which encode the proteins represented
 CC in AAV73814-Y74252.
 CC
 SQ Sequence 1890 BP; 590 A; 358 C; 398 G; 544 T; 0 other;
 Query Match 56.0%; Score 14; DB 20; Length 1890;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 caagtcacagatt 24
 Db 1475 CAAGTCACAGATT 1462
 RESULT 15
 AAV10120/c
 ID AAV10120 standard; cDNA; 2233 BP.
 XX
 AC AAV10120;
 XX
 XX 29-MAY-1998 (first entry)
 DT XX
 DE Human retinoid receptor RRL cDNA.
 XX
 KM Retinoid receptor; RRL; steroid receptor; agonist; antagonist; cancer;
 KM adrenal deficiency; skin disorder; inflammatory disorder;
 KW immune response regulator; autoimmune disease; therapeutic antibody; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 363..1781
 FT /*tag= a
 FT /product= RRL
 FT /note= "retinoid receptor"
 XX
 XX US5728548-A.
 XX
 PD 17-MAR-1998.
 PD XX
 PF 29-JUN-1995; 95US-0496631.
 PF XX
 XX 29-JUN-1995; 95US-0496631.
 PR XX
 XX (GENY) GENETICS INST INC.
 PA XX
 PI Bowman M;
 XX

DR WPI; 1998-206567/18.
 DR P-PSDB; AAM40072.
 XX
 PT Human retinoid receptor protein RRL - useful for, e.g. drug
 PT screening, therapy and antibody production
 XX
 XX Claim 8; Column 11-15; 13pp; English.
 PS
 XX
 CC This sequence encodes a novel human steroid receptor, the retinoid
 CC receptor protein or RRL. This protein can be used in screening assays
 CC for steroid hormone receptor agonists and antagonists and in
 CC pharmaceutical compositions for treating adrenal deficiencies, e.g.
 CC Addison's disease, cancer, skin disorders, e.g. acne and psoriasis,
 CC inflammatory disorders, e.g. arthritis and HIV infections. The protein
 CC can also be used for regulating immune responses, e.g. as antitumor
 CC agents, vaccine adjuvants, organ rejection inhibitors or agents for
 CC treating autoimmune diseases. The protein can further be used to produce
 CC therapeutic antibodies.
 CC
 SQ Sequence 2233 BP; 744 A; 427 C; 463 G; 599 T; 0 other;
 Query Match 56.0%; Score 14; DB 19; Length 2233;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 ttgtcgaagtcac 18
 Db 1055 TTGTCGCAAGTCAC 1042
 RESULT 16
 AAF77541/c
 ID AAF77541 standard; DNA; 2585 BP.
 XX
 AC AAF77541;
 XX
 XX 23-MAY-2001 (first entry)
 DT XX
 DE Caenorhabditis briggsae mab-21 coding sequence.
 XX
 KM Mab-21; cell fate choice; pax-6; aniridia; Moebius syndrome;
 KM chromosome 13q13; peripheral neural tissue differentiation; ds.
 XX
 OS Caenorhabditis briggsae.
 XX
 PN US6197504-B1.
 PN XX
 PD 06-MAR-2001.
 PD XX
 PF 19-JAN-1998; 98US-0008697.
 PF XX
 XX 10-APR-1996; 96US-0011607.
 PR 09-APR-1997; 97US-0835604.
 PR XX
 XX (CHOW/) CHOW K L.
 PA XX
 XX Chow KL;
 PI XX
 XX WPI; 2001-256362/26.
 DR P-PSDB; AAB74419.
 DR XX
 XX
 PT Detecting expression of mab-21 gene of Caenorhabditis elegans encoding
 PT novel protein required for choice of alternate cell fates in sample, by
 PT contacting mRNA from sample with nucleic acid hybridizing with mab-21 -
 XX
 PS Disclosure; column 41-44; 79pp; English.
 XX
 CC The present invention describes a method of detecting mab-21 expression
 CC in a sample involving contacting total mRNA from the sample with a probe
 CC to the mab-21 gene. The sequences of the mab-21 coding sequence and
 CC protein from several species are also given. The human mab-21 gene is
 CC found on chromosome 13q13, a region which has been linked to Moebius

CC syndrome. The mab-21 protein shows homology to pax-6, mutations in which
CC are associated with aniridia and which is linked to peripheral neural
CC tissue differentiation.
XX
SQ Sequence 2585 BP; 751 A; 540 C; 559 G; 688 T; 47 other;

Query Match 56.0%; Score 14; DB 22; Length 2585;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 caagtcacagaatt 24
|||||
DB 1058 CAGGTCCAGCAATT 1045

RESULT 17
AAQ1851/c
ID AAQ1851 standard; DNA; 3083 BP.
XX
AC AAQ1851;
XX
DT 31-JUL-1991 (first entry)
XX
DE Glutamate receptor 3.
XX
KW Glutamate receptor 3; probe; ligand; drug screening; ss.
XX
OS Rattus rattus.
XX
FH Key location/Qualifiers
FT CDS 167..2830
FT /*tag= a
FT /product= GR3
FT sig_peptide 167..232
FT /*tag= b
FT mat_peptide 233..2830
FT /*tag= c
XX
PN WC09106648-A.
XX
PD 16-MAY-1991.
XX
PF 25-OCT-1990; 90WO-US06153.
XX
PR 27-OCT-1989; 89US-0428116.
XX
PA (SALK) SALK INST FOR BIOL. STUD.
XX
PI Heinemann SF, Boulter JR, Hollmann M, Bettler B, Jensen JE;
XX
DR WPI; 1991-164197/22.
XX
DR P-PSDB; AAR11991.
XX
PT Glutamate receptors - used to screen for functional ligands and
XX identify and isolate further receptors
XX
PS Disclosure; Fig 4; 109pp; English.
XX
CC GUR3 cDNA was isolated from a rat forebrain cDNA using a
CC low-stringency screening protocol and a radiolabelled fragment of the
CC GUR1 cDNA as probe. The cDNA is deposited (ATCC 68133).
CC The gene and protein can be used in drug screening, to
CC determine whether a substance is a functional ligand for the
CC receptor by monitoring ion channel activity.
CC See also AAQ1849-855.
XX
SQ Sequence 3083 BP; 877 A; 679 C; 735 G; 792 T; 0 other;

Query Match 56.0%; Score 14; DB 12; Length 3083;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 catgtgcgaagtc 16
|||||
DB 3034 CATGTGCAGCAATC 3021

RESULT 18
AAC76014/c
ID AAC76014 standard; CDNA; 5773 BP.
XX
AC AAC76014;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORF1569 polynucleotide sequence SEQ ID NO:3137.

Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antineoplastic;
KW antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antineoplastic disease; coagulation;
KW thrombosis; contraceptive; ss.

Homo sapiens.

WO200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000WO-US08621.

31-MAR-1989; 99US-0127607.

02-APR-1999; 99US-0127636.

05-APR-1999; 99US-0127728.

30-MAR-2000; 2000US-0540763.

(CURA-) CURAGEN CORP.

Shinkets RA, Leach M;

WPI; 2000-602362/57.

P-PSDB; AAB41805.

Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 2354-2357; 5507pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antineoplastic; antibacterial; antiviral; antifungal; antirheumatic;
CC antihypertensive; antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus, lupus
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus

CC erythematous, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC neonatal haemoglobinuria, anti-inflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 SO Sequence 5773 BP; 1691 A; 1125 C; 1269 G; 1687 T; 1 other;

Query Match 56.0%; Score 14; DB 21; Length 5773;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gaagtcacagat 23
 |||||
 Db 4611 GCAGTCCACAGAT 4598

RESULT 19
 AAF21307
 ID AAF21307 standard; DNA: 32351 BP.
 AC AAF21307;
 XX
 DT 14-MAR-2001 (first entry)
 XX
 DE Human low adenosine antisense oligonucleotide related sequence #2874.
 XX
 KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
 KW human; airway disorder; bronchoconstriction; lung inflammation;
 KW surfactant depletion; respiratory; bronchodilator; anti-inflammatory;
 KW immunosuppressive; antihistaminic; analgesic; hypotensive; cytostatic;
 KW respiratory obstruction; pulmonary obstruction; impeded respiration;
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
 KW cancer; ss.
 KW
 XX
 OS Homo sapiens.
 XX
 PN WO200062736-A2.
 XX
 PD 26-OCT-2000.
 XX
 PF 24-MAR-2000; 2000WO-US08020.
 XX
 PR 06-APR-1999; 99US-0127958.
 XX
 PA (UYEC-) UNIV EAST CAROLINA.
 PA (NYCE/) NYCE J W.
 PI
 XX
 NYce JW;
 XX
 DR WPI: 2000-679539/66.
 XX
 PT Low adenosine (A) content antisense oligonucleotides which do not
 PT trigger adenosine receptors during metabolism, useful e.g. for treating
 PT cancers and respiratory obstructions -
 XX
 PS Disclosure: Page 1295-1303; 1592pp; English.

CC The present invention describes low adenosine (A) content antisense
 CC oligonucleotides and compositions (I) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
 CC (I) can have respiratory, bronchodilator, anti-inflammatory, analgesic,
 CC immunosuppressive, antihistaminic, hypotensive and cytostatic activities.
 CC The antisense oligonucleotides and (I) can be used to down-regulate the
 CC expression and/or activity of target polypeptides associated with
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC

CC binding proteins, adhesion molecules and their receptors, cytokine and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS) and peripheral nervous and non-nervous system
 CC receptors, CNS and peripheral nervous and non-nervous system peptide
 CC transmitters, defensins, growth factors, vasoactive peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
 CC and/or surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impeded respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.
 CC
 SO Sequence 32351 BP; 8594 A; 7026 C; 7405 G; 9326 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 32351;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aagtcacagatta 25
 |||||
 Db 10696 aagtcacagatta 10709

RESULT 20
 AAA35185
 ID AAA35185 standard; DNA: 32351 BP.
 AC AAA35185;
 XX
 DT 28-JUL-2000 (first entry)
 XX
 DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:59.
 XX
 KW Human: adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphorothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; anti-inflammatory;
 KW antiallergic; antihistaminic; cytostatic; analgesic; impaired airway;
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
 KW
 XX
 OS Homo sapiens.
 XX
 PN WO200009525-A2.
 XX
 PD 24-FEB-2000.
 XX
 PF 03-AUG-1999; 99WO-US17712.
 XX
 PR 03-AUG-1998; 98US-0095212.
 XX
 PA (UYEC-) UNIV EAST CAROLINA.
 PA
 PI
 XX
 NYce JW;
 XX
 DR WPI: 2000-205971/18.
 XX
 PT New antisense oligonucleotides useful for treating e.g. pulmonary
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
 PT cancers -
 XX
 PS Disclosure: Page 1212-1219; 1343pp; English.

CC The present invention describes a new composition comprising an
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
 CC inflammation. The ON can have antiinflammatory, antiallergic,
 CC antisthmatic, cytosstatic and analgesic activities. The compositions are
 CC useful for the treatment of diseases associated with inflammation,
 CC impaired airways, including lung disease and diseases whose secondary
 CC effects afflict the lungs of a subject. They can be used for treating
 CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
 CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
 CC carcinomas, and cancers which may metastasise to the lungs, including
 CC breast and prostate cancer. The reduction of the adenosine content of
 CC the ONs reduces side effects. The A-containing ONs break down with the
 CC release of deoxyadenosine which activates adenosine receptors causing
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
 CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680
 CC (AAA33323 to AAA33992) are specifically claimed ONs from the present
 CC invention. N.B. Sequences given in the disclosure of the present
 CC invention do not match up with their corresponding SEQ ID NO: sequences
 CC given in the sequence listing.

XX
 XX Sequence 32351 BP; 8594 A; 7027 C; 7405 G; 9325 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 32351;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aagtcacagaatta 25
 |||||
 DB 10696 aagtcacagaatta 10709

RESULT 21
 AAF21311
 ID AAF21311 standard; DNA: 40298 BP.

XX
 AC AAF21311;
 DT 14-MAR-2001 (first entry)

XX Human low adenosine antisense oligonucleotide related sequence #2878.

DE
 XX Low adenosine antisense oligonucleotide: phosphorothioate; allergy;
 KW human; airway disorder; bronchoconstriction; lung inflammation;
 KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
 KW immunosuppressive; antisthmatic; analgesic; hypotensive; cytosstatic;
 KW respiratory obstruction; pulmonary obstruction; impaired respiration;
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
 KW cancer; ss.

XX Homo sapiens.

XX WO200062736-A2.

XX 26-OCT-2000.

XX 24-MAR-2000; 2000WO-US08020.

XX 06-APR-1999; 99US-0127958.

XX (UYEC-) UNIV EAST CAROLINA.

XX (NYCE/) NYCE J W.

XX NYCE JW;
 XX PN

XX
 DR WPI: 2000-679539/66.

XX Low adenosine (A) content antisense oligonucleotides which do not
 PT trigger adenosine receptors during metabolism, useful e.g. for treating
 PT cancers and respiratory obstructions -

XX Disclosure: Page 1305-1315; 1592pp; English.

XX The present invention describes low adenosine (A) content antisense
 CC oligonucleotides and compositions (1) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'universal' or alternative base.
 CC (1) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antisthmatic, hypotensive and cytosstatic activities.
 CC The antisense oligonucleotides and (1) can be used to down-regulate the
 CC expression and or activity of target polypeptides associated with
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokine and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS) and peripheral nervous and non-nervous system
 CC receptors, CNS and peripheral nervous and non-nervous system peptide
 CC transmitters, defensins, growth factors, vasoactive peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
 CC and/or surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impaired respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.

XX Sequence 40298 BP; 10485 A; 9119 C; 9484 G; 11210 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 40298;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aagtcacagaatta 25
 |||||
 DB 11521 aagtcacagaatta 11534

RESULT 22
 AAA35189
 ID AAA35189 standard; DNA: 40298 BP.

XX AAA35189;

XX 28-JUL-2000 (first entry)

XX Human adenosine receptor related polynucleotide 2nd SEQ ID NO:63.

DE
 XX Human; adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphorothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antisthmatic; cytosstatic; analgesic; impaired airway;
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

XX Homo sapiens.

XX WO200009525-A2.

XX 24-FEB-2000.
 XX 03-AUG-1999; 99WO-US17712.
 XX 03-AUG-1998; 98US-0095212.
 XX (UYEC-) UNIV EAST CAROLINA.
 XX NYCE JW;
 XX WPI: 2000-205971/18.
 XX
 XX New antisense oligonucleotides useful for treating e.g. pulmonary
 XX vasoconstriction, inflammation, allergies, asthma, hypertension,
 XX bronchitis, emphysema, respiratory distress syndrome, ischemia or
 XX cancers -
 XX
 XX Disclosure: Page 1221-1231; 1343pp; English.
 XX
 XX The present invention describes a new composition comprising an
 XX antisense oligonucleotide (ON) with low adenosine (up to 15%), which
 XX targets nucleic acids involved in bronchoconstriction, allergies, and/or
 XX inflammation. The ON can have anti-inflammatory, antiallergic,
 XX antisthmatic, cyostatic and analgesic activities. The compositions are
 XX useful for the treatment of diseases associated with inflammation,
 XX impaired airways, including lung disease and diseases whose secondary
 XX effects afflict the lungs of a subject. They can be used for treating
 XX e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
 XX asthma, impaired respiration, respiratory distress syndrome, pain, cystic
 XX fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 XX pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
 XX carcinomas, and cancers which may metastasise to the lungs, including
 XX breast and prostate cancer. The reduction of the adenosine content of
 XX the ONs reduces side effects. The A-containing ONs break down with the
 XX release of deoxyadenosine which activates adenosine receptors causing the
 XX bronchoconstriction and inflammation. AAA3213 to AAA3532 represent the
 XX nucleotide sequences given in the sequence listing from the present
 XX invention, which correspond to SEQ ID NO:1 to 2815, and then the last
 XX 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
 XX differ from the previously named sequences. SEQ ID NO:11 to 1680
 XX (AAA3223 to AAA3392) are specifically claimed ONs from the present
 XX invention. N.B. Sequences given in the disclosure of the present
 XX invention do not match up with their corresponding SEQ ID NO: sequences
 XX given in the sequence listing.
 XX
 XX Sequence 40298 BP; 10485 A; 9121 C; 9482 G; 11210 T; 0 other;
 SQ
 Query Match 56.0%; Score 14; DB 21; Length 40298;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 12 aagtcacagaatta 25
 ||||||||||||
 DB 11521 aagtcacagaatta 11534
 RESULT 23
 AA201425
 ID AA201425 standard; DNA; 1038602 BP.
 XX
 XX AA201425;
 XX
 XX 07-OCT-1999 (first entry)
 XX
 XX Complete genome sequence of Chlamydia trachomatis.
 DE
 XX
 XX Vaccine; eye disease; conventional trachoma; nongendemic trachoma;
 KW Paratrachoma; inclusion conjunctivitis; genital disease; peritrapatitis;
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 KW Bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.
 XX
 XX

OS Chlamydia trachomatis.
 XX
 XX WO9928475-A2.
 XX
 XX 10-JUN-1999.
 XX
 XX 27-NOV-1998; 98WO-IB01939.
 XX
 XX 04-NOV-1998; 98US-0107077.
 XX 28-NOV-1997; 97FR-0015041.
 XX 17-DEC-1997; 97FR-0016034.
 XX
 XX (GEST) GENSET.
 XX
 XX Griffiths R;
 XX
 XX WPI: 1999-371125/31.
 XX
 XX Genome sequence of Chlamydia trachomatis
 XX
 XX Claim 1; Page 373-656; 1755pp; English.
 XX
 XX The present sequence represents the complete genome of Chlamydia
 XX trachomatis. Open reading frames (ORFs) of the genome encode
 XX polypeptides AA36754-Y37949. The polypeptides can be used as vaccines
 XX against Chlamydia trachomatis. Antisense and ribozyme sequences can also
 XX be used to control growth of the microorganism. Chlamydia trachomatis is
 XX responsible for a large number of diseases, e.g. eye diseases such as
 XX conjunctivitis; genital diseases such as nongonococcal urethritis;
 XX epididymitis; cervicitis; salpingitis; peritrapatitis; Bartholinitis;
 XX pneumopathy in breast feeding infants; and venereal
 XX lymphogranulomatosis. The polypeptides of the invention may be of use in
 XX treating these diseases.
 XX
 XX Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;
 SQ
 Query Match 56.0%; Score 14; DB 20; Length 1038602;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 attgtcgcaagta 17
 ||||||||||||
 DB 736214 attgtcgcaagta 736227
 RESULT 24
 AAT22576
 ID AAT22576 standard; cDNA to mRNA; 87 BP.
 XX
 XX AAT22576;
 XX
 XX 01-OCT-1996 (first entry)
 XX
 XX Human gene signature HUMGS04192.
 DE
 XX
 XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 KW cell typing; abnormal cell function; ss.
 XX
 XX Homo sapiens.
 XX
 XX WO9514772-A1.
 XX
 XX 01-JUN-1995.
 XX
 XX 11-NOV-1994; 94WO-JP01916.
 XX
 XX 12-NOV-1993; 93JP-0355504.
 XX
 XX (MATS/) MATSUBARA K.
 XX (OKUB/) OKUBO K.
 XX
 XX

```

XX Matsubara K, Okubo K;
PI
XX
XX WPI; 1995-206931/27.
DR
XX
XX Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 1160; 2245bp; Japanese.
XX
XX A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in AAT19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
XX
XX Sequence 87 BP; 31 A; 25 C; 9 G; 21 T; 1 other;
SQ

Query Match          52.0%; Score 13; DB 16; Length 87;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aagtcacagaatt 24
   |||||
DB 16 aagtcacagaatt 28

RESULT 25
AAT20233/C
ID AAT20233 standard; cDNA to mRNA; 167 BP.
XX
XX AAT20233;
AC
XX
XX 24-JUL-1996 (first entry)
DT
XX
XX Human gene signature HMGSO1380.
DE
XX
XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
XX
XX Homo sapiens.
OS
XX
XX MO9514772-A1.
PN
XX
XX 01-JUN-1995.
PD
XX
XX 11-NOV-1994; 94MO-JP01916.
PF
XX
XX 12-NOV-1993; 93JP-0355504.
PR
XX
XX (MATS/) MATSUBARA K.
PA (OKUBO/) OKUBO K.
XX
XX Matsubara K, Okubo K;
PI
XX
XX WPI; 1995-206931/27.
DR
XX
XX Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that

```

```

PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PS Claim 1; Page 592; 2245bp; Japanese.
XX
XX A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in AAT19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
XX
XX Sequence 167 BP; 65 A; 23 C; 26 G; 44 T; 9 other;
SQ

Query Match          52.0%; Score 13; DB 16; Length 167;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 agtcacagaatta 25
   |||||
DB 118 AGTCACAGATT 106

RESULT 26
AAC26213/C
ID AAC26213 standard; cDNA; 187 BP.
XX
XX AAC26213;
AC
XX
XX 06-OCT-2000 (first entry)
DT
XX
XX Human secreted protein 5' EST, SEQ ID NO: 30288.
DE
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
OS
XX
XX EP1033401-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 21-FEB-2000; 2000EP-0200610.
PF
XX
XX 26-FEB-1999; 99US-0122487.
PR
XX
XX (GEST ) GENSET.
PA
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI
XX
XX WPI; 2000-500381/45.
DR
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 30288; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or poly(A) RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)

```

CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

XX Sequence 187 BP; 59 A; 27 C; 30 G; 71 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 187;

Best Local Similarity 100.0%; Pred. No. 75;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aagtcacagaatt 24
 |||
 Db 163 AAGTCACAGAATT 151

RESULT 27

AAA50159/c
 ID AAA50159 standard; cDNA; 363 BP.

XX AAA50159;

DT 07-NOV-2000 (first entry)

XX Antibody 5H7 heavy chain variable region cDNA.

XX 5H7: monoclonal antibody; apoptosis; programmed cell death;

KM major histocompatibility complex class I; MHC; tumour; therapy;

KM graft-mediated immunoprotection; single chain antibody; scfv; ss.

XX Mammalia.

PN WO200047713-A2.

XX 17-ANG-2000.

XX 08-FEB-2000; 2000WO-US03234.

XX 09-FEB-1999; 99US-0119238.

XX (ARCH-) ARCH DEV CORP.

PI Woodie ES, Van Seventer JM, Kulkarni S, Kranz D, Holman P;

XX WPI: 2000-558210/51.

DR P-PSDB; AAY95776.

XX New polynucleotides encoding single chain light and heavy variable

PT regions of a 5H7 antibody, useful in methods for conferring programmed

PT cell death and graft-mediated immune protection of cell, tissue and

PT organs -

XX Claim 2; Page 36; 48pp; English.

XX The present sequence is that of cDNA coding for the heavy chain
 CC variable region (VH) (see AAY95776) of 5H7, an anti-human class I
 CC major histocompatibility complex (MHC) monoclonal antibody (MAb)
 CC which recognises a monomorphic determinant of the alpha-3 domain.
 CC MAb 5H7 has been shown to induce programmed cell death (PCD) in
 CC lymphoid tumours, peripheral blood mononuclear cells, and B6 mouse
 CC splenocytes transgenic for class I HLA-B27. Polynucleotides
 CC encoding 5H7 VL and VH sequences were used in the construction of
 CC single chain variable immunoglobulin domains (scfv) of the 5H7
 CC antibody (see AAY95781). These can be used to confer PCD properties
 CC to cells, especially T and B lymphocyte tumour cells. Recombinant
 CC DNA molecules encoding 5H7 scfv are also used in methods for
 CC conferring graft-mediated immune protection of cell, tissue and

CC transplanted organs.

XX Sequence 363 BP; 88 A; 101 C; 84 G; 90 T; 0 other;

SQ Query Match 52.0%; Score 13; DB 21; Length 363;

Best Local Similarity 100.0%; Pred. No. 75;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aagtcacagaatt 24
 |||
 Db 262 AAGTCACAGAATT 250

RESULT 28
 AAQ25667/c
 ID AAQ25667 standard; cDNA; 378 BP.

XX AAQ25667;

DT 28-DEC-1992 (first entry)

XX Sequence of the anti-urokinase antibody kappa variable region (VK)

DE cDNA.

KM Chimeric monoclonal antibody; anti-urokinase antibody; PCR;

XX antithrombotic agent; myocardial infarction therapy; ss.

XX Mus musculus.

XX Key

FT sig_peptide 1..9 location/Qualifiers

FT mat_peptide 10..378 /tag= a

FT misc_feature 1..27 /tag= b

FT misc_feature 18..32 /tag= c

FT misc_feature 340..371 /tag= d

FT misc_feature 356..371 /tag= e

FT misc_feature 356..371 /tag= f

FT misc_feature 356..371 /tag= g

FT misc_feature 356..371 /tag= h

FT misc_feature 356..371 /tag= i

FT misc_feature 356..371 /tag= j

FT misc_feature 356..371 /tag= k

FT misc_feature 356..371 /tag= l

FT misc_feature 356..371 /tag= m

FT misc_feature 356..371 /tag= n

FT misc_feature 356..371 /tag= o

FT misc_feature 356..371 /tag= p

FT misc_feature 356..371 /tag= q

FT misc_feature 356..371 /tag= r

FT misc_feature 356..371 /tag= s

FT misc_feature 356..371 /tag= t

FT misc_feature 356..371 /tag= u

CC the mc-kappa primer as primer for first strand synthesis, and the
 CC 3'mv-kappa and 5'mv-kappa primers as primers for the PCR. The
 CC amplified fragment was restriction digested and ligated into a
 CC restriction fragment of pTB1423 to give an anti-tyrosinase antibody
 CC vK cDNA confg. Plasmid pTB1456. The sequence of this plasmid is
 CC given in AA025667. The cDNA is a functional vK gene. The sequence of
 CC the primer (mc-gamma-1) is given in AA025689.

SO Sequence 378 BP; 92 A; 100 C; 91 G; 95 T; 0 other;

Query Match

Best Local Similarity 52.0%; Score 13; DB 13; Length 378;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aagtcacagaatt 24
 |||
 DB 271 AAGTCACAGAAAT 259

RESULT 29

AAA74604/c
 ID AAA74604 standard; cDNA; 408 BP.

XX AAA74604;

XX 08-JAN-2001 (first entry)

DE Murline 5B3 antibody heavy chain variable region cDNA.

XX Mouse; 5B3 antibody; IgG1; piezoelectric immunoassay;

KM small molecule explosive detection; 2,4,6-trinitrotoluene; TNT; ss.

XX Mus sp.

OS Location/Qualifiers

Key 1..408

FT CDS /*tag- a

FT /partial

FT /product= "5B3 antibody heavy chain variable region"

PN WO200043774-A2.

XX 27-JUL-2000.

XX 25-JAN-2000; 2000WO-IL00048.

XX 25-JAN-1999; 99IL-0128212.

XX (YISS) YISSUM RES & DEV CO.

PA (YEDA) YEDA RES & DEV CO LTD.

PI Willner I, Eshhar Z;

DR WPI; 2000-524259/47.

DR P-PSDB; AAB15672.

PT Apparatus for detecting small molecules, especially explosives

PT comprises a piezoelectric crystal .

PS Disclosure; Fig 3A; 90pp; English.

XX The present sequence encodes the heavy chain variable region of 5B3
 CC antibody. It was obtained from total RNA extracted from a 5B3 hybridoma
 CC by RT-PCR. 5B3 is an IgG1 antibody derived from a TNP-KiH immunised
 CC mouse. Its binding to TNP or DNS antigen can be blocked by very low
 CC amounts of TNT and it can therefore be used in a method for detecting
 CC small assayed explosive molecules. Molecules are detected using a
 CC piezoelectric sensor. Piezoelectric immunoassaying in liquid phase
 CC allows stationary and flow analysis of an aqueous sample. The method
 CC is sufficiently sensitive for detection of low molecular weight
 CC molecules.

SO Sequence 408 BP; 90 A; 108 C; 99 G; 111 T; 0 other;

Query Match

Best Local Similarity 52.0%; Score 13; DB 21; Length 408;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 aagtcacagaatt 24
 |||
 DB 316 AAGTCACAGAAAT 304

RESULT 30

AAc56783/c
 ID AAC56783 standard; DNA; 415 BP.

XX AAC56783;

XX 25-JAN-2001 (first entry)

DE Eucalyptus grandis transcription factor DNA sequence #654.

XX plant; transcription factor; gene expression; eucalyptus; pine; acacia;

KM poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;

KM basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;

KM homeodomain zipper; LIM domain; AP2; ERBS; zinc finger domain;

XX type 2 Cys2His2; CCAAT box element; MYB; ss.

OS Eucalyptus grandis.

XX WO200053724-A2.

XX 14-SEP-2000.

XX 09-MAR-2000; 2000WO-US06112.

XX 11-MAR-1999; 99US-0266513.

XX 18-AUG-1999; 99US-0149485.

XX (GENE-) GENESIS RES & DEV CORP LTD.

XX (FLEET-) FLETCHEER CHALLENGE FORESTS LTD.

XX Wood M, McGrath A, Shenk MA, Glenn M;

XX WPI; 2000-579369/54.

XX Claim 1; Pages 513-514; 747pp; English.

XX The present invention relates to novel plant transcription factors from
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
 CC sequence for one such transcription factor. The transcription factor may
 CC be used to produce a plant having modified gene expression such as a
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
 CC mahogany species or to modify the activity of a polypeptide in a plant.
 CC The transcription factors of the present invention are members from the
 CC following families of regulatory proteins: bZIP, bZIP family of G-box
 CC binding factors, basic helix-loop-helix zipper,
 CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
 CC and ERBS, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 CC and MYB.

SO Sequence 415 BP; 124 A; 99 C; 88 G; 104 T; 0 other;

Query Match

Best Local Similarity 52.0%; Score 13; DB 21; Length 415;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 atgtcgcagatc 16
 |||||
 DB 121 ATTGTCGCAAGTC 109

RESULT 31
 AAC09392/c
 ID AAC09392 standard; cDNA; 476 BP.
 XX
 AC AAC09392;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 13467.

KM Human: 5' EST: expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PS (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 1: SEQ ID 13467; 71pp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 XX

SEQ Sequence 476 BP; 109 A; 72 C; 90 G; 200 T; 5 other;

Query Match 52.0%; Score 13; DB 21; Length 476;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caagtcacagaat 23
 |||||
 DB 263 CAAGTCACAGAAAT 251

RESULT 32
 AAF93531/c
 ID AAF93531 standard; cDNA; 559 BP.
 XX
 AC AAF93531;
 XX

DT 21-MAY-2001 (first entry)

DE cDNA encoding SRT protein isolated from testis tissue SEQ ID 352.

KM Human: SRT; gene therapy; gene mapping; tissue typing; ss.

OS Homo sapiens.

PN WO200107611-A2.

PD 01-FEB-2001.

PF 21-JUL-2000; 2000WO-US20006.

PR 26-JUL-1999; 99US-0145701.

PS (GETH) GENENTECH INC.

PI Baker KP, Goddard A, Wood WT;

DR WPI; 2001-112729/12.

XX New isolated nucleic acid molecule encoding a SRT polypeptide is useful
 PT for production of recombinant SRT polypeptides, gene mapping,
 PT diagnosing genetic disorders and for gene therapy -
 XX
 PS Claim 2: Fig 352; 663pp; English.

CC Sequences AAF93180 - AAF93743 represent polynucleotide sequences encoding
 CC human SRT proteins. The cDNA sequences are isolated from various
 CC different human tissue libraries. The invention relates to a method
 CC for detecting cDNA encoding an SRT protein, a vector containing cDNA
 CC encoding SRT, a host cell transformed with the vector, an isolated SRT
 CC polypeptide, and an antibody which binds to SRT. The polynucleotide
 CC sequence can be used in gene therapy and is useful in the recombinant
 CC production of SRT polypeptides, as a hybridisation probe to screen
 CC libraries to isolate cDNAs with sequence identity to SRT polypeptides, to
 CC map the gene encoding the SRT polypeptides and analysing genetic
 CC disorders, tissue typing and disease tissue detection. The SRT
 CC polynucleotide sequences can be used in polymerase chain reaction,
 CC screening for new therapeutic molecules and generation of antisense RNA
 CC and DNA.
 XX

SEQ Sequence 559 BP; 127 A; 127 C; 130 G; 171 T; 4 other;

Query Match 52.0%; Score 13; DB 22; Length 559;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caagtcacagaat 23
 |||||
 DB 78 CAAGTCACAGAAAT 66

RESULT 33
 AAA95830
 ID AAA95830 standard; cDNA; 637 BP.

AC AAA95830;

DT 23-FEB-2001 (first entry)

DE Bovine metalloproteinase KIAA0688 cDNA.

XX Bovine: KIAA0688; metalloproteinase; ADAM;
 KW a disintegrin and metalloproteinase domain; thrombospondin domain;
 KW vaccine; neutropenic; neuroprotective; antiparkinsonian;
 KW cerebroprotective; cytosolic; antiarthritic; immunosuppressive;
 KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
 KW autoimmune disease; brain tumour; brain injury; ss.
 XX
 OS Bos taurus.


```

XX Key Location/Qualifiers
FH CDS 2..370
FT //tag= a
FT //partial
FT //product= "K1AA0688"
XX
XX W0200053774-A2.
XX
XX PD 14-SEP-2000.
XX
XX PF 08-MAR-2000; 2000WO-US06237.
XX
XX PR 08-MAR-1999; 99US-0264585.
XX
XX PA (NEUR-) NEUROCRINE BIOSCIENCES INC.
XX
XX PI Keiner GS, Clark M, Maki RA;
XX
XX DR WPI; 2000-594326/56.
XX
XX DR P-PSDB; AAB21260.
XX
XX PT Polynucleotide encoding novel members of a disintegrin,
XX metalloproteinase and thrombospondin domain protein family used to
XX prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
XX
XX PS Example 3; Fig 20; 129pp; English.
XX
XX CC The present sequence encodes bovine metalloproteinase KIAA0688, which is
XX a member of the ADAMTS family of proteins. Members of the ADAMTS family
XX contain a metalloproteinase, integrin and thrombospondin domain. ADAMTS
XX polypeptides are useful for the manufacture of medicaments for treating
XX conditions associated with neuroinflammation and/or neurodegeneration,
XX such as Alzheimer's disease, Parkinson's disease and stroke. They are
XX also useful for treating conditions associated with cell proliferation,
XX cell migration, inflammation and/or angiogenesis, such as cancer,
XX arthritis and autoimmune diseases. They can be used to treat patients
XX afflicted with an invasive tumour, a brain tumour or brain injury.
XX
XX SQ Sequence 637 BP; 107 A; 227 C; 196 G; 107 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 637;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 cgcaagtcacagaa 21
Db 314 cgcaagtcacagaa 326

RESULT 34
AAQ81957/c
ID AAQ81957 standard; cDNA to mRNA; 713 BP.
XX
XX AC AAQ81957;
XX
XX DT 15-MAR-1995 (first entry)
XX
XX DE C2P2(487-713).
XX
XX KW Canine; dog; zona pellucida; ZP; C2P2; contraceptive; vaccine;
XX antigen; ss.
XX
XX OS Canis familiaris.
XX
XX PN JRP6189766-A.
XX
XX PD 12-JUL-1994.
XX
XX PF 25-DEC-1992; 92JP-0359265.
XX
XX PR 25-DEC-1992; 92JP-0359265.

```

```

XX PA (TOFU) TONEN CORP.
XX
XX DR WPI; 1994-259553/32.
XX
XX PT New DNA sequence encoding canine zona pellucida C2P2 - useful for
XX the prodn. of a canine contraceptive vaccine antigen
XX
XX PS Disclosure; Page 6; 10pp; Japanese.
XX
XX CC The C2P2 DNA (AAQ70072) was prep'd. by the cloning of C2P2(75-520) -
XX AAQ81700 using the primers given in AAQ70073-74, C2P2(1-65) - AAQ81804
XX using the primers given in AAQ70082-83, C2P2(42-103) - AAQ81803 using
XX the primers given in AAQ70079-81 and C2P2(487-713) - AAQ81957 using the
XX primers given in AAQ70075-76.
XX
XX SQ Sequence 713 BP; 177 A; 189 C; 162 G; 185 T; 0 other;

Query Match 52.0%; Score 13; DB 15; Length 713;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacagaa 22
Db 429 GCAAGTCACAGAA 417

RESULT 35
AAT79355/c
ID AAT79355 standard; DNA; 777 BP.
XX
XX AC AAT79355;
XX
XX DT 23-MAR-1998 (first entry)
XX
XX DE Staphylococcus aureus downstream junction sequence LO2C4.
XX
XX KW Staphylococcus aureus; junction; upstream; downstream; mec domain;
XX methicillin resistance; coagulase negative; ds.
XX
XX OS Staphylococcus aureus.
XX
XX PN W09731125-A2.
XX
XX PD 28-AUG-1997.
XX
XX PF 21-FEB-1997; 97WO-JP00487.
XX
XX PR 23-FEB-1996; 96JP-0060373.
XX
XX PA (HIRA/) HIRAMATSU K.
XX (KATN-) KAINOS LAB INC.
XX
XX PI Aways A, Hayashi T, Ohno H;
XX
XX DR WPI; 1997-435174/40.
XX
XX PT Diagnosing methicillin resistant Staphylococcus aureus or
XX coagulase-negative Staphylococcus - by detecting junction sequences
XX at upstream and downstream points where methicillin resistance
XX domain adjoins rest of chromosome
XX
XX PS Claim 5; Fig 12; 71pp; Japanese.
XX
XX CC This LO2C4 is the downstream junction sequence of a Staphylococcus
XX aureus DNA. This is used in a method for diagnosing methicillin resistant
XX Staphylococcus aureus (MRSa) or methicillin resistant coagulase-negative
XX Staphylococcus (MRC-NS). The method comprises extracting nucleic acid
XX molecules from a Staphylococci containing sample, and detecting the
XX junction sequences at the upstream and downstream points where the
XX methicillin resistance (mec) domain adjoins the rest of the chromosome.
XX The sequences are subjected to NASBA-dot hybridisation. The method can be

```

CC used for the detection of MRSA and MRC-NS in biological samples such as
 CC blood, urine and sputum, and their discrimination from methicillin-
 CC sensitive strains.

XX Sequence 777 BP; 249 A; 130 C; 91 G; 307 T; 0 other;

Query Match 52.0%; Score 13; DB 18; Length 777;
 Best Local Similarity 100.0%; Pred. No. 76;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 agtcacagaatta 25
 |||||
 DB 636 AGTCACAGAATTA 624

RESULT 36

AA32456/C
 ID AAX32456 standard; DNA; 778 BP.

AC AAX32456;

DF 22-JUN-1999 (first entry)

DE Partial nucleotide sequence of NCTC10442.

KW Methicillin-resistant coagulase-negative Staphylococci; MRC-NS; LCR;
 KW meca; hybridisation; MSC-NS; MRSA; methicillin-resistant S. aureus;
 KW PCR; NASBA; ss.

OS Staphylococcus sp.

PN JP11056371-A.

PD 02-MAR-1999.

PF 22-AUG-1997; 97JP-0242070.

PR 22-AUG-1997; 97JP-0242070.

PA (HIRA/) HIRAMATSU K.

PA (ITOK/) ITO K.

PA (KAIN-) KAINOSU KK.

DR WPI: 1999-222384/19.

PT Identification of methicillin-resistant Staphylococcus aureus (MRSA)
 PT or methicillin-resistant coagulase-negative Staphylococci (MRC-NS) -
 PT useful for rapid analysis of biological samples

PS Examples; Fig 15-16; 45pp; Japanese.

CC The invention relates to a method for the identification of methicillin-
 CC resistant coagulase-negative Staphylococci (MRC-NS) in which part of
 CC mecDNA which is a foreign inserted DNA present on the chromosome of MRC-
 CC NS and carrying meca gene and part of the base sequence of the chromosome
 CC DNA surrounding the inserted DNA are used together to react with a sample
 CC specimen. The method identifies MRC-NS by NASBA, PCR, LCR or
 CC hybridisation by utilising a fact that the base sequence of a chromosome
 CC DNA surrounding the inserted site of mecDNA in the chromosome of
 CC methicillin sensitive coagulase negative Staphylococci (MSC-NS) is used
 CC for a reaction with a sample and when mecDNA is inserted it comes
 CC negative. The method can be used for especially identifying methicillin-
 CC resistant Staphylococcus aureus (MRSA). The method can detect MRSA from a
 CC specimen such as urine, blood, sputum, etc. in 30 minutes to 1 hour.

SO Sequence 778 BP; 249 A; 130 C; 91 G; 308 T; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 778;
 Best Local Similarity 100.0%; Pred. No. 76;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 agtcacagaatta 25
 |||||
 DB 637 AGTCACAGAATTA 625

RESULT 37

AA256756/C
 ID AA256756 standard; cDNA; 902 BP.

AC AA256756;

DF 23-MAR-2000 (first entry)

DE Human transmembrane protein HTPN-59 encoding cDNA.

KW Human; transmembrane protein; HTPN; diagnosis; immunospecific;
 KW antiproliferative; neuroprotective; immune disorder;
 KW reproductive disorder; smooth muscle disorder; neurological disorder;
 KW gastrointestinal disorder; developmental disorder;
 KW cell proliferative disorder; ss.

OS Homo sapiens.

PN WO9961471-A2.

PD 02-DEC-1999.

PF 28-MAY-1999; 99WO-US11904.

PR 29-MAY-1998; 98US-0087260.

PR 02-JUL-1998; 98US-0091674.

PR 02-OCT-1998; 98US-0102854.

PR 24-NOV-1998; 98US-0109869.

PA (INCY-) INCYTE PHARM INC.

PI Tang YT, Lal P, Hillman JL, Yue H, Guegler KJ, Corley NC;
 PI Bandman O, Patterson C, Gorgone GA, Kaser MR, Baughn MR;
 PI Au-Young J;

DR WPI: 2000-072605/06.

DR P-PDB; AA57935.

PT Proteins, polynucleotides, vectors, host cells and antibodies used to
 PT diagnose, treat or prevent immune, reproductive, smooth muscle,
 PT neurological, gastrointestinal, developmental and cell proliferative
 PT disorders -

PS Claim 9; Page 215; 229pp; English.

CC AA256698 to AA256776 encode AA57877 to AA57955 which represent human
 CC transmembrane proteins designated HTPN-1 to HTPN-79, respectively.
 CC The transmembrane protein have immunospecific, antiproliferative and
 CC neuroprotective activities. The human transmembrane proteins,
 CC polynucleotides encoding them and other compositions and methods from
 CC the present invention, can be used for the diagnosis, treatment or
 CC prevention of immune, reproductive, smooth muscle, neurological,
 CC gastrointestinal, developmental and cell proliferative disorders. The
 CC HTPN's can be used to treat or prevent disorders associated with a
 CC decreased expression or activity of HTPN.

SO Sequence 902 BP; 261 A; 212 C; 196 G; 233 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 902;
 Best Local Similarity 100.0%; Pred. No. 76;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 agtcacagaatta 25
 |||||
 DB 425 AGTCACAGAATTA 413

RESULT 38
AAAI5335/c
ID AAII5335 standard; cDNA; 930 BP.
XX
XX
AC AAII5335;
XX
DT 20-MAY-1999 (first entry)
XX
XX cDNA encoding a lectin recognising glucose.
DE
XX
XX Lectin protein; glucose; maboya; fungal disease; Candidosis;
KM Aspergillus; Cryptococcus; mucormycosis; trichosporonosis; ss.
XX
OS Halocynthia roretzi.
XX
XX Key Location/Qualifiers
FT CDS 52..726
FT /*tag= a
XX
XX JP11032759-A.
XX
XX 09-FEB-1999.
XX
XX 17-JUL-1997; 97JP-0208580.
XX
XX 17-JUL-1997; 97JP-0208580.
XX
XX (FUJI/) FUJITA T.
XX
XX WPI; 1999-183820/16.
XX P-PSDB; AAW97115.
XX
XX A new lectin protein and DNA - useful for prevention and treatment
PT of fungal diseases such as Candidosis, Aspergillus,
PT Cryptococcus, mucormycosis and trichosporonosis
XX
XX Disclosure; Fig 1; 9pp; Japanese.
XX
XX The present sequence encodes a lectin protein that recognises
CC glucose. The protein is isolated from maboya body fluid. The protein
CC is useful in agents for prevention and treatment of fungal diseases
CC such as Candidosis, Aspergillus, Cryptococcus, mucormycosis and
CC trichosporonosis.
XX
XX Sequence 930 BP; 320 A; 177 C; 203 G; 230 T; 0 other;
SQ

Query Match 52.0%; Score 13; DB 20; Length 930;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 4 attgtcgaagtc 16
|||
Db 437 ATTGTGCGCACTC 425

RESULT 39
AAA50164/c
ID AAA50164 standard; DNA; 1002 BP.
XX
XX
AC AAA50164;
XX
XX 07-NOV-2000 (first entry)
XX
XX DNA encoding 5H7 single chain antibody (scfv).
DE
XX
XX 5H7; monoclonal antibody; apoptosis; programmed cell death;
KM major histocompatibility complex class I; MHC; tumour; therapy;
KM graft-mediated immunoprotection; scfv; single chain antibody; ss.
XX
XX Chimeric - mammalia.
OS
XX Chimeric - synthetic.

PN W0200047713-A2.
XX
XX 17-AUG-2000.
XX
XX 08-FEB-2000; 2000WO-US03234.
XX
XX 09-FEB-1999; 99US-0119238.
XX
XX (ARCH-) ARCH DEV CORP.
XX
XX Woodlee ES, Van Seventer JM, Kulkarni S, Kranz D, Holman P;
PI WPI; 2000-558210/51.
XX
XX P-PSDB; AAV95781.
XX
XX New polynucleotides encoding single chain light and heavy variable
PT regions of a 5H7 antibody, useful in methods for conferring programmed
PT cell death and graft-mediated immune protection of cell, tissue and
PT organs -
XX
XX Disclosure; Page 41-42; 48pp; English.
XX
XX The present sequence is that of a recombinant DNA molecule encoding
CC a single chain antibody (scfv). The scfv is based on 5H7, an
CC anti-human class I major histocompatibility complex (MHC) monoclonal
CC antibody (MAB) that confers programmed cell death (PCD) inducing
CC properties to cells. The scfv is composed of (from the N-terminus)
CC a leader peptide, the light chain variable region of MAB 5H7, a
CC linker peptide, the heavy chain variable region of MAB 5H7, a
CC second linker peptide, and glycosylphosphatidylinositol. The
CC recombinant DNA molecule can be placed into a variety of expression
CC systems and hosts for production of the scfv. A method for
CC conferring PCD inducing properties to cells involves transforming a
CC cell with the recombinant DNA molecule. The method provides for PCD
CC signalling in B and T lymphocyte tumour cells. In a related
CC method, graft-mediated immune protection is conferred to a cell,
CC tissue or organ (especially an organ for transplant) by
CC transformation with the recombinant DNA molecule.
XX
XX Sequence 1002 BP; 258 A; 258 C; 237 G; 249 T; 0 other;
SQ

Query Match 52.0%; Score 13; DB 21; Length 1002;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 12 aagtcacagaatt 24
|||
Db 733 AAGTCACAGAAATT 721

RESULT 40
AAC64778
ID AAC64778 standard; cDNA; 1010 BP.
XX
XX
AC AAC64778;
XX
XX 28-FEB-2001 (first entry)
XX
XX Strawberry alcohol dehydrogenase encoding cDNA SEQ ID NO:13A.
DE
XX
XX Strawberry; fruit flavour; biosynthetic pathway; aliphatic; thiolase;
KM aromatic ester; alcohol acyl transferase; alcohol dehydrogenase;
KM pyruvate decarboxylase; aminotransferase; esterase; alcohol; aldehyde;
KM alpha-keto acid; amino acid; fatty acid; acyl-CoA; processed food;
KM food additive; flavour; syrup; ice-cream; frozen dessert; yoghurt;
KM confectionery; flavouring; oral medication; vitamin; aroma; beverage;
KM alcohol; scent; fragrance; perfume; cosmetic; suspension aid;
KM aluminium salt; anti-perispirant; pharmaceutical; cleaning product;
KM insect pheromone; dye carrier; solvent; insect repellent; miticide;
KM scabicide; plasticiser; deodorant; ss.
XX
XX Fragaria x ananassa.

```

XX  WO200032789-A1.
PN
XX
XX  08-JUN-2000.
PD
XX
XX  02-DEC-1999; 99WO-NL00737.
PF
XX
XX  02-DEC-1998; 98EP-0204018.
PR
XX  12-MAR-1999; 99EP-0200739.
PR
XX  (CPRO-) CPRO-DIO CENT PLANTENVEREDELINGS REPROD.
PA
XX  Aharoni A, Luecker J, Verhoeven HA, Van Tunen AJ, O'Connell AP;
PI
XX  WPI: 2000-412335/35.
DR
XX  P-PSDB: AAB36451.
DR
XX
XX  A new DNA sequence encoding a polypeptide with alcohol acyl transferase
PT  activity for producing and regulating aromatic and/or aliphatic ester
PT  formation in microorganisms, plant cells or plants -
XX
XX  Claim 37; Page 100-101; 163pp; English.
PS
XX
XX  The present invention describes nucleotide sequences with thiolase,
CC  alcohol acyl transferase, alcohol dehydrogenase, pyruvate decarboxylase,
CC  aminotransferase and esterase activities, which are involved in the
CC  biosynthetic pathway for aliphatic and/or aromatic ester production in
CC  fruit. The nucleotide sequences can be inserted into the genome of a
CC  fruit-producing plant to regulate aliphatic and/or aromatic ester
CC  formation. Aromatic and/or aliphatic esters in microorganisms, plant
CC  cells or plants are produced by inserting thiolase, alcohol acyl
CC  transferase, alcohol dehydrogenase, pyruvate decarboxylase,
CC  aminotransferase and esterase nucleotide sequences into the genome and
CC  feeding the microorganism or plant with alcohol, aldehydes, alpha-keto
CC  acids, or amino acids and fatty acids, and acyl-CoA. The nucleotides and
CC  their proteins can be used in the processed food industry as food
CC  additives to enhance the flavour of syrups, ice-creams, frozen desserts,
CC  yoghurts and confectionery. They are used: as flavouring agents for oral
CC  medications and vitamins; provide flavour and aroma in beverages,
CC  including alcohol; enhance or reduce fruit flavour, aroma, fragrance or
CC  scent; enhance the flavour or aroma of natural, synthetic or artificial
CC  products; for the production of novel combinations of artificial flavour
CC  substances; as antibacterial or anti-fungal agents; as fragrance or
CC  perfumes in cosmetics, creams, sun-protectant products, hair
CC  conditioners, lengthening agents and fixatives in perfumes, suspension
CC  aids for aluminium salts in anti-perspirant pharmaceuticals, cleaning
CC  products, personal care products and animal care products; as
CC  disinfectant additives; as degreasing solvents for electronics; as
CC  insect pheromones; and as dye carriers, solvents, insect repellents,
CC  miticides, scabicides, plasticisers and deodorants. The present sequence
CC  encodes the specifically claimed strawberry alcohol dehydrogenase.
XX
XX  Sequence 1010 BP; 333 A; 171 C; 237 G; 269 T; 0 other;
SQ

```

```

Query Match      52.0%; Score 13; DB 21; Length 1010;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY  10 gcaagtcacagaa 22
    |||||
DB  385 gcaagtcacagaa 397

```

```

RESULT 41
ID  AAA27678
XX  AAA27678 standard; cDNA; 1010 BP.
AC  AAA27678;
XX
XX  29-AUG-2000 (first entry)
DT
XX  Strawberry-alcohol-dehydrogenase partial cDNA clone SLG144.
DE

```

```

XX  Strawberry: alcohol dehydrogenase; fruit; ripening; ester;
XX  flavour; aroma; transgenic plant; ss.
KW
XX
XX  Fragaria ananassa.
OS
XX
XX  Key Location/Qualifiers
FH  2..853
FT  CDS /*tag= a
FT  /partial
XX
XX  EP1006190-A1.
PN
XX
XX  07-JUN-2000.
PD
XX
XX  02-DEC-1998; 98EP-0204018.
PF
XX
XX  02-DEC-1998; 98EP-0204018.
PR
XX  (CPRO-) CPRO-DIO CENT PLANTENVEREDELINGS REPROD.
PA
XX  Verhoeven HA, van Tunen AJ, Aharoni A, Luecker J, O'Connell AP;
PI
XX  WPI: 2000-378264/33.
DR
XX  P-PSDB: AAY79668.
DR
XX
XX  New polynucleotides encoding enzymes from the biosynthetic pathway for
PT  aromatic and/or aliphatic ester production in fruit used to modify
PT  plant flavours -
XX
XX  Example 3; Page 93-94; 116pp; English.
PS
XX
XX  The present sequence is that of a partial cDNA clone for strawberry
CC  cv. Elsanta alcohol dehydrogenase SLG144 (see AAY79668), an enzyme
CC  that shows upregulated expression in ripening fruit. The clone was
CC  isolated from a cDNA library by expression analysis using Northern
CC  blotting. The invention relates to DNA sequences (see AAA27666-78)
CC  encoding enzymes (see AAY29656-68) involved in the metabolic pathway
CC  leading to the formation of aliphatic and/or aromatic esters in
CC  ripening fruit. The enzymes have alcohol acyl transferase,
CC  alcohol dehydrogenase, pyruvate decarboxylase, thiolase or
CC  aminotransferase activity. Expression vectors comprising the DNA
CC  sequences may be used to regulate ester formation in fruit.
CC  Genetically modified plants, plant cells and microorganisms can be
CC  used to produce esters. The DNA sequences, polypeptides and
CC  antibodies are also used to screen fruit; for volatile ester
CC  compounds; for quality such as flavour, fragrance, aroma, scent,
CC  texture or shape; to distinguish between cultivars and varieties;
CC  and to monitor harvest time, post-harvest quality, shelf-life,
CC  timing of pesticide application, and resistance capacity based on
CC  volatile ester profiles.
XX
XX  Sequence 1010 BP; 333 A; 171 C; 237 G; 269 T; 0 other;
SQ

```

```

Query Match      52.0%; Score 13; DB 21; Length 1010;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY  10 gcaagtcacagaa 22
    |||||
DB  385 gcaagtcacagaa 397

```

```

RESULT 42
ID  AAC36039/C
XX  AAC36039 standard; DNA; 1278 BP.
AC  AAC36039;
XX
XX  17-OCT-2000 (first entry)
DT
XX  Arabidopsis thaliana DNA fragment SPQ ID NO: 12322.
DE

```

XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134376.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137529.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144003.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145224.
PR 23-JUL-1999; 99US-0145228.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.

```

PR 31-AUG-1999; 99US-0151438-
PR 01-SEP-1999; 99US-0151930-
PR 07-SEP-1999; 99US-0152363-
PR 10-SEP-1999; 99US-0153070-
PR 13-SEP-1999; 99US-0153758-
PR 15-SEP-1999; 99US-0154018-
PR 16-SEP-1999; 99US-0154039-
PR 20-SEP-1999; 99US-0154779-
PR 22-SEP-1999; 99US-0155139-
PR 23-SEP-1999; 99US-0155486-
PR 24-SEP-1999; 99US-0155639-
PR 28-SEP-1999; 99US-0156458-
PR 29-SEP-1999; 99US-0156596-
PR 04-OCT-1999; 99US-0157117-
PR 05-OCT-1999; 99US-0157753-
PR 06-OCT-1999; 99US-0157865-
PR 07-OCT-1999; 99US-0158029-
PR 08-OCT-1999; 99US-0158232-
PR 12-OCT-1999; 99US-0158369-
PR 13-OCT-1999; 99US-0159293-
PR 13-OCT-1999; 99US-0159294-
PR 13-OCT-1999; 99US-0159295-
PR 14-OCT-1999; 99US-0159329-
PR 14-OCT-1999; 99US-0159330-
PR 14-OCT-1999; 99US-0159331-
PR 14-OCT-1999; 99US-0159637-
PR 14-OCT-1999; 99US-0159638-
PR 18-OCT-1999; 99US-0159584-
PR 21-OCT-1999; 99US-0160741-
PR 21-OCT-1999; 99US-0160767-
PR 21-OCT-1999; 99US-0160770-
PR 21-OCT-1999; 99US-0160814-
PR 21-OCT-1999; 99US-0160815-
PR 22-OCT-1999; 99US-0160980-
PR 22-OCT-1999; 99US-0160981-
PR 22-OCT-1999; 99US-0160989-
PR 22-OCT-1999; 99US-0161404-
PR 25-OCT-1999; 99US-0161405-
PR 25-OCT-1999; 99US-0161406-
PR 26-OCT-1999; 99US-0161359-
PR 26-OCT-1999; 99US-0161360-
PR 26-OCT-1999; 99US-0161361-
PR 28-OCT-1999; 99US-0161920-
PR 28-OCT-1999; 99US-0161921-
PR 28-OCT-1999; 99US-0161992-
PR 28-OCT-1999; 99US-0161993-
PR 29-OCT-1999; 99US-0162142-

```

```

Query Match 52.0%; Score 13; DB 21; Length 1278;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 3 caltgcgcaagt 15
Db 78 CATTGTCCAGT 66

```

```

RESULT 43
AAA94840/c
ID AAA94840 standard; DNA; 1317 BP.

```

```

AC AAA94840;

```

```

XX 23-FEB-2001 (first entry)

```

```

DE Drosophila melanogaster odorant receptor DOR44 coding sequence.

```

```

XX Odorant receptor; fruit fly; DOR44; odour recognition; pest control; ds.

```

```

OS Drosophila melanogaster.

```

```

XX Key Location/Qualifiers
FH CDS 1..1317
FT

```

```

FT /*tag= a
FT /product= "DOR44"
FT /partial

```

```

W0200050566-A2.

```

```

31-AUG-2000.

```

```

XX 25-FEB-2000; 2000WO-US04995.

```

```

XX 25-FEB-1999; 99US-0257706.

```

```

XX (UYCO ) UNIV COLUMBIA NEW YORK.

```

```

XX Vosshall LB, Amrein HO, Axel R;

```

```

DR WPI: 2000-572081/53.

```

```

XX P-PSDB; AAB26441.

```

```

PT Novel nucleic acid encoding an insect odorant receptor, for identifying
modulator compounds that are useful in controlling pest population

```

```

XX Disclosure; Page 70-71; 176pp; English.

```

```

CC The present sequence is the coding sequence for the previously identified
CC Drosophila melanogaster odorant receptor DOR44. The odorant genes and
CC proteins, such as those provided by the invention, are useful as they aid
CC in the study of the olfactory organ in mammals, as well as aiding the
CC understanding of the link between odour recognition and behaviour in
CC insects. They also enable the identification of compounds capable of
CC activating and inhibiting the receptors, allow the control of pest
CC populations via the use of alarm odour ligands and via the use of ligands
CC which interfere with the interaction between odorant ligands and
CC receptors associated with fertility.

```

```

XX SQ Sequence 1317 BP; 299 A; 317 C; 327 G; 374 T; 0 other;

```

```

Query Match 52.0%; Score 13; DB 21; Length 1317;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 11 caagtcacagaat 23
Db 162 CAGTCCACAGAT 150

```

```

RESULT 44
AAA72281/c
ID AAA72281 standard; cDNA; 1359 BP.

```

```

AC AAA72281;

```

```

XX 06-DEC-2000 (first entry)

```

```

DE Drosophila odorant receptor DORU 25.1 cDNA.

```

```

XX Odorant receptor; Drosophila; olfactory receptor;

```

```

XX G protein-coupled receptor; GPCR superfamily; transgenic insect;
XX insect behaviour modification; pest control; pollinator attraction;
XX biosensor; odour detection; odour identification; apiculture; ss.

```

```

OS Drosophila melanogaster.

```

```

XX W0200043410-A2.

```

```

XX 27-JUL-2000.

```

```

XX 25-JAN-2000; 2000WO-US01823.

```

```

XX 25-JAN-1999; 99US-0117132.

```

```

XX (UYTA ) UNIV YALE.

```

XX Carlson JR, Kim J, Clyne PJ, Warr CG;
 XX WPI: 2000-543246/49.
 DR P-PSDB; AAB20946.
 XX
 PT New nucleic acid encoding a *Drosophila* olfactory receptor, useful for
 PT identifying modulating agents -
 XX
 XX Claim 3; Page 271-273; 303pp; English.
 XX
 CC Sequences AAA72236-A72284 represent cDNAs encoding *Drosophila*
 CC melanogaster odorant receptors (AAB20901-B20949). These proteins function
 CC as olfactory receptors, and are thought to be members of the G
 CC protein-coupled receptor (GPCR) superfamily, which is characterised by
 CC the presence of 7 transmembrane helices. Nucleic acids encoding the
 CC *Drosophila* odorant receptors may be used to generate expression
 CC constructs, host cells containing such constructs, and transgenic
 CC insects. Cells which express the odorant receptor genes may be used in
 CC methods to identify agents which modulate expression of these genes, and
 CC in methods to identify receptor binding partners. The *Drosophila* odorant
 CC receptor nucleic acids may also be used to identify corresponding genes
 CC in other insects, such as those which damage crops or transmit disease.
 CC The odorant receptor proteins may be used to identify agents which
 CC modulate their activity, to identify binding partners, as antigens to
 CC raise antibodies, and in methods to modify insect behaviour. The proteins
 CC may be also be used in methods of behaviour modification. Such methods
 CC may be used to study or modify insect behaviour in response to odorants
 CC such as pheromones. Modification of insect behaviour has a wide range of
 CC applications, such as in pest control (e.g., by disrupting the feeding
 CC or mating behaviours of pest species), or for enhancing plant
 CC pollination (by attracting pollinator species). Odorant receptor proteins
 CC and/or nucleotides may also be used to identify appetite suppressants, to
 CC trap odours of a specific type, as biosensors for the detection of
 CC explosives, drugs, perfumes or pollutants, and in apiculture to modify
 CC the behaviour of bees, for example, to increase the production of royal
 CC jelly.
 CC
 XX Sequence 1359 BP; 312 A; 329 C; 333 G; 385 T; 0 other;
 SO

Query Match 52.0%; Score 13; DB 21; Length 1359;
 Best Local Similarity 100.0%; Pred. No. 76;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caagtcacagaat 23
 ||||||||||||
 Db 162 CAAGTCACAGAAAT 150

RESULT 45
 AA259716/c
 ID AA259716 standard; DNA: 1404 BP.
 XX
 AC AA259716;
 XX
 DT 19-APR-2000 (first entry)
 XX
 DE DNA encoding a mutant phytase-10, phytase-10-thermo[3]-Q50T-K91A.
 XX
 KM Phytase: myo-inositol hexakisphosphate phosphohydrolase; stabilisation;
 KM thermostable; animal feed; monogastric animal; phytate phosphorus;
 KM phosphate availability; consensus; mutant; ds.
 XX
 XX Aspergillus terreus 9A1.
 OS Aspergillus terreus cbs16.46.
 OS Aspergillus niger var. awamori.
 OS Aspergillus niger str. NRRL3135.
 OS Aspergillus fumigatus ATCC13073.
 OS Aspergillus fumigatus ATCC32722.
 OS Aspergillus fumigatus ATCC58128.
 OS Aspergillus fumigatus ATCC26906.
 OS Aspergillus fumigatus ATCC32239.

OS *Emicella nidulans*.
 OS *Talaromyces thermophilus* ATCC20186.
 OS *Myxiliophthora thermophila*.
 OS *Paxillus involutus* NN005693.
 OS *Trametes pubescens* NN9343.
 OS *Agrocycbe pedicades* NN009289.
 OS *Peniophora lycii* NN006113.
 OS *Thermomyces lanuginosa*.
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FT CDS 1..1404
 FT /*tag= a
 FT /product= "Phytase-10-thermo[3]-Q50T-K91A"
 XX
 XX EP669089-A1.
 PD 05-JAN-2000.
 XX
 PF 23-JUN-1999; 99EP-0111949.
 XX
 XX 29-JUN-1998; 98EP-0111960.
 PR
 XX (HOFF) HOFFMANN IA ROCHE & CO AG F.
 PA
 PI Brugger R, Lehmann M, Wyss M;
 XX
 XX WPI: 2000-099429/09.
 DR
 XX
 PT New stabilised enzyme formulation, useful for feed compositions for
 PT monogastric animals -
 PT
 XX Example 5; Fig 19; 101pp; English.
 XX
 CC The invention relates to a novel stabilised dry or liquid enzyme
 CC formulation, comprising phytase (myo-inositol hexakisphosphate
 CC phosphohydrolase) and one or more stabilising agents including
 CC xylyl or ribitol; polyethylene glycols with a molecular weight of 600
 CC to 4000 Da, preferably 1000 to 3350 Da; the disodium salts of malonic,
 CC glutaric and succinic acid; carboxymethylcellulose; and sodium alginate.
 CC The stabilised phytase formulation is used in a method for preparing a
 CC feed composition for monogastric animals (e.g., pigs, poultry) and
 CC provides a monogastric animal with its dietary requirements of
 CC phosphorus. Although a large amount of phosphate is present in animal
 CC feed in the form of phytate phosphorus, monogastric animals are unable
 CC to utilise this form of phosphate, resulting in the addition of extra
 CC phosphate to the feed of such animals. Phytase enhances the nutritional
 CC value of plant material without the need for adding additional phosphate
 CC to the feed. The level of phosphate pollution in the environment is
 CC reduced by adding phytase to animal feed, as the animal can make use of
 CC the inorganic phosphate liberated from phytate phosphorus using the
 CC enzyme. The phytase formulation of the invention has an improved
 CC thermostability and can therefore remain stable during long-term storage
 CC and can withstand feed processing methods such as extrusion, expansion
 CC and pelleting. The present sequence represents DNA encoding a mutant
 CC phytase-10 consensus sequence, phytase-10-thermo[3]-Q50T-K91A, which has
 CC a temperature optimum and melting point 4 degrees Celsius higher than
 CC that of phytase-10 (AA69566). Its specific activity with phytate as a
 CC substrate is also strongly increased.
 XX
 SO Sequence 1404 BP; 329 A; 309 C; 308 G; 458 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 1404;
 Best Local Similarity 100.0%; Pred. No. 76;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caagtcacagaat 23
 ||||||||||||
 Db 852 CAAGTCACAGAAAT 840

Wed Oct 10 07:45:31 2001

us-09-396-196f-1.oli.rng

Page 24

Search completed: October 9, 2001, 15:51:43
Job time: 15064 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 12:15:24 ; Search time 218.82 Seconds
(Without alignments)
21.629 Million cell updates/sec

Title: US-09-396-196F-1

Sequence: 1 gacatgtcgcagtcacagaa125

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 324599 seqs, 9465562 residues

Word size: 9

Total number of hits satisfying chosen parameters: 7928

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: Issued Patents, NA:*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCRTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-401-068-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-1
5	17	68.0	17	2	US-07-944-259-2
6	15	60.0	1771	2	US-09-166-203-48
7	14	56.0	1050	6	5498529-7
8	14	56.0	1677	2	US-08-372-652-14
9	14	56.0	1677	2	US-08-372-652-14
10	14	56.0	1787	2	US-08-372-652-6
11	14	56.0	1787	2	US-08-372-652-6
12	14	56.0	2233	1	US-08-496-631-1
13	14	56.0	2585	4	US-09-008-697A-7
14	14	56.0	3083	1	US-07-718-575-5
15	14	56.0	3083	1	US-08-481-206-5
16	14	56.0	3083	2	US-08-486-269A-5
17	14	56.0	1662	1	US-08-651-572-1
18	14	56.0	1662	3	US-09-066-544-1
19	13	52.0	1662	3	US-08-951-086-1
20	13	52.0	2091	1	US-08-765-081-4
21	13	52.0	2091	5	US-09-098-082-4
22	13	52.0	2091	5	PCT-US95-06994-4
23	13	52.0	2214	2	US-08-484-993B-1
24	13	52.0	2214	2	US-08-484-158B-1
25	13	52.0	2214	2	US-08-484-596A-1
26	13	52.0	2214	2	US-08-480-150A-1
27	13	52.0	2214	3	US-08-458-731-1

C 28	13	52.0	2214	3	US-08-149-223A-1	Sequence 1, Appl
C 29	13	52.0	2381	2	US-08-484-993B-9	Sequence 9, Appl
C 30	13	52.0	2381	2	US-08-484-158B-9	Sequence 9, Appl
C 31	13	52.0	2381	2	US-08-484-596A-9	Sequence 9, Appl
C 32	13	52.0	2381	2	US-08-480-150A-9	Sequence 9, Appl
C 33	13	52.0	2381	2	US-08-458-731-9	Sequence 9, Appl
C 34	13	52.0	2381	3	US-08-149-223A-9	Sequence 9, Appl
C 35	13	52.0	4500	2	US-08-743-637B-35	Sequence 35, Appl
C 36	13	52.0	4500	3	US-08-526-840B-35	Sequence 35, Appl
C 37	13	52.0	8041	1	US-08-765-081-1	Sequence 1, Appl
C 38	13	52.0	8041	3	US-09-098-082-1	Sequence 1, Appl
C 39	13	52.0	8041	5	PCT-US95-06994-1	Sequence 1, Appl
C 40	13	52.0	72928	3	US-09-009-913-1	Sequence 1, Appl
C 41	12	48.0	24	3	US-09-014-065-15	Sequence 15, Appl
C 42	12	48.0	255	5	PCT-US93-03077-7	Sequence 7, Appl
C 43	12	48.0	279	1	US-08-047-041A-17	Sequence 17, Appl
C 44	12	48.0	308	4	US-09-232-522-3	Sequence 3, Appl
C 45	12	48.0	311	1	US-08-047-041A-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-401-068-7
Sequence 7, Application US/08401068
Patent No. 5859335
GENERAL INFORMATION:
APPLICANT: Patton, David
TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESS: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/401,068
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/351,970
FILING DATE: 08-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1038
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /product= "Biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 5098
OTHER INFORMATION: /function= "unknown", involved in biotin synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ORF1"
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "promoter plac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name= "promoter plac"
FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name= "bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name= "bioA RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name= "ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name= "rho-independent
OTHER INFORMATION: /transcriptional terminator"
FEATURE:
NAME/KEY: stem loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "promoter plac"
OTHER INFORMATION: /evidence= EXPERIMENTAL
PUBLICATION INFORMATION: /number= 2
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 5.2e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gacattgtcgcagtcacagaaatla 25
|||||
Db 137 GACATTGTCTCGCAGTCACAGAAATTA 161

RESULT 4
US-08-411-768B-6
Sequence 6, Application US/08411768B
Patent No. 6083712
GENERAL INFORMATION:
APPLICANT: Olwen Birch
APPLICANT: Johann Brass

APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pRO30A15-9
FEATURE:
NAME/KEY: CDS
LOCATION: 1154..2308
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 1154
OTHER INFORMATION: /EC_number= 2.3.1.47
OTHER INFORMATION: /product= "KAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioF"
OTHER INFORMATION: /number= 2
OTHER INFORMATION: /standard_name= "8-Amino-7-oxononanoate synthase"
FEATURE:
NAME/KEY: CDS
LOCATION: 3043..3753
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3043
OTHER INFORMATION: /EC_number= 6.3.3.3
OTHER INFORMATION: /product= "DTPB synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioD"
OTHER INFORMATION: /number= 4
OTHER INFORMATION: /standard_name= "Delhiobiotin synthase"
FEATURE:
NAME/KEY: RBS
LOCATION: 1141..1156
OTHER INFORMATION: /standard_name= "bioF RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3030..3045
OTHER INFORMATION: /standard_name= "biob RBS"
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1

FLING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-7688-6

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 5.2e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacattgtcgaagtcacagaatta 25
|||||
DB 137 GACATTGTGCAAGTCACAGAAATTA 161

RESULT 5

US-07-944-259-2/C
Sequence 2, Application US/07944259
Patent No. 5885792

GENERAL INFORMATION:

APPLICANT: Ifuku, Ohji
APPLICANT: Haze, Shintaro
APPLICANT: Kishimoto, Jiro
APPLICANT: Nakahama, Kazuo
TITLE OF INVENTION: BIOTIN OPERON
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wegner, Cantor, Mueller & Player
STREET: 1233 20th Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036-8218

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/944,259
FILING DATE: 19920914
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Cantor, Herbert I.
REGISTRATION NUMBER: 24,392
REFERENCE/DOCKET NUMBER: P-450-23557
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-887-0400
TELEFAX: 202-835-0605
TELEX: 440706

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 17 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
US-07-944-259-2

Query Match 68.0%; Score 17; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gtgcgaagtcacagaat 23
|||||
DB 17 GTGCGAAGTCACAGAAAT 1

RESULT 6
US-09-166-203-48
Sequence 48, Application US/09166203A

Patent No. 5968826
GENERAL INFORMATION:
APPLICANT: Bennett, C. Frank
APPLICANT: Condon, Tom P.
APPLICANT: Cowsett, Lex M.
TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION
FILE REFERENCE: ISPH-0323
CURRENT APPLICATION NUMBER: US/09/166,203A
CURRENT FILING DATE: 1998-10-05
NUMBER OF SEQ ID NOS: 60
SEQ ID NO 48

LENGTH: 1771
TYPE: DNA
ORGANISM: Mus musculus

FEATURE:

NAME/KEY: CDS
LOCATION: (1193)..(1387)
FEATURE:
NAME/KEY: CDS
LOCATION: (1709)..(1771)
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: L20788 Genbank
DATABASE ENTRY DATE: 1996-04-18
US-09-166-203-48

Query Match 60.0%; Score 15; DB 2; Length 1771;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacagaatt 24
|||||
DB 312 gcaagtcacagaatt 326

RESULT 7

5498529-7/C
Patent No. 5498529

APPLICANT: BERKA, THOMAS R.; FORMWALD, JAMES A.; GORNIAK,
JOSEPHIN G.; ROSENBERG, MARTIN; STRICKLER, JAMES E.; TAYLOR, DEAN P.

TITLE OF INVENTION: PROTEIN PROTEASE INHIBITORS FROM
STREPTOMYCES

NUMBER OF SEQUENCES: 9

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/162,506
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 894,167
FILING DATE: 03-JUN-1992
APPLICATION NUMBER: 346,119
FILING DATE: 16-FEB-1989
APPLICATION NUMBER: 897,245
FILING DATE: 18-AUG-1986
SEQ ID NO: 7
LENGTH: 1050
5498529-7

Query Match 56.0%; Score 14; DB 6; Length 1050;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aagtcacagaatta 25
|||||
DB 251 AAGTCACAGAAATTA 238

RESULT 8
US-08-372-652-14/C
Sequence 14, Application US/08372652
Patent No. 5932699
GENERAL INFORMATION:
APPLICANT: Moore, David

APPLICANT: Seol, Mongi
FILING DATE: 13-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/246001
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1677 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
PCT-US95-16311-14

Query Match 56.0%; Score 14; DB 2; Length 1677;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 ttgtcgcaagtcac 18
|||||

Db 630 TTGTCCCACTCAGTCA 617

RESULT 9
PCT-US95-16311-14/c
Sequence 14, Application PC/TUS9516311
GENERAL INFORMATION:
APPLICANT: Moore, David
APPLICANT: Seol, Mongi
TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
TITLE OF INVENTION: POLYPEPTIDES AND RELATED MOLECULES AND METHODS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street, Suite 3100
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16311
FILING DATE:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/372,652
FILING DATE: 13-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/246001
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1677 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
PCT-US95-16311-14

Query Match 56.0%; Score 14; DB 5; Length 1677;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 ttgtcgcaagtcac 18
|||||

Db 630 TTGTCCCACTCAGTCA 617

RESULT 10
US-08-372-652-6/c
Sequence 6, Application US/08372652
Patent No. 5932699
GENERAL INFORMATION:
APPLICANT: Moore, David
APPLICANT: Seol, Mongi
TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
TITLE OF INVENTION: POLYPEPTIDES AND RELATED MOLECULES AND METHODS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street, Suite 3100
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/372,652
FILING DATE: 13-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/246001
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1787 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-372-652-6

Query Match 56.0%; Score 14; DB 2; Length 1787;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ttgtcgcaagtcac 18
|||||
DB 740 TTGTGCAAGTCAC 727

RESULT 11
PCT-US95-16311-6/c
Sequence 6, Application PC/TUS9516311
GENERAL INFORMATION:
APPLICANT: Moore, David
APPLICANT: Choi, Hwang-Sik
TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street, Suite 3100
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16311
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/372,652
FILING DATE: 13-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/246001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1787 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
PCT-US95-16311-6

Query Match 56.0%; Score 14; DB 5; Length 1787;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ttgtcgcaagtcac 18
|||||
DB 740 TTGTGCAAGTCAC 727

RESULT 12
US-08-496-631-1/c
Sequence 1, Application US/08496631
Patent No. 5728548
GENERAL INFORMATION:
APPLICANT: Bowman, Michael
TITLE OF INVENTION: STEROID RECEPTOR RRI
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/496,631
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15248
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2233 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 363..1778
US-08-496-631-1

Query Match 56.0%; Score 14; DB 1; Length 2233;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ttgtcgcaagtcac 18
|||||
DB 1055 TTGTGCAAGTCAC 1042

RESULT 13
US-09-008-697A-7/c
Sequence 7, Application US/09008697A
Patent No. 6197504
GENERAL INFORMATION:
APPLICANT: Chow, King Lau
TITLE OF INVENTION: USES OF MAB-21
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/008,697A
FILING DATE: January 19, 1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Chan, Albert Wai-Kit
REGISTRATION NUMBER: 36,479
REFERENCE/DOCKET NUMBER: 50752-A

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2585 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-008-697A-7

Query Match 56.0%; Score 14; DB 4; Length 2585;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 caatgcacagaatt 24
|||||
DB 1058 CAATGCACAGATT 1045

RESULT 14
US-07-718-575-5/C
Sequence 5, Application US/07718575
Patent No. 5202257
GENERAL INFORMATION:
APPLICANT: Heinemann Ph.D., Stephen F.
APPLICANT: Boulter Ph.D., James R.
APPLICANT: Hollmann Ph.D., Michael NMN
APPLICANT: Bettler Ph.D., Bernhard NMN
APPLICANT: Jensen Ph.D., Jan E.
TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND
METHODS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
STREET: 444 So. Flower St., Suite 2000
CITY: Los Angeles
STATE: California
COUNTRY: United States
ZIP: 90071-2921
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/718,575
FILING DATE: 19910813
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Reiter Ph.D., Stephen E.
REGISTRATION NUMBER: 31192
REFERENCE/DOCKET NUMBER: P31 8962
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
TELEX: 9103330318
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: G1UR3
FEATURE:
NAME/KEY: CDS
LOCATION: 167..2833
US-07-718-575-5

Query Match 56.0%; Score 14; DB 1; Length 3083;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 catgtcgcaagtc 16
|||||
DB 3034 CATGTGCGCAAGTC 3021

RESULT 15
US-08-481-206-5/C
Sequence 5, Application US/08481206
Patent No. 5739291
GENERAL INFORMATION:
APPLICANT: Heinemann Ph.D., Stephen F.
APPLICANT: Boulter Ph.D., James R.
APPLICANT: Hollmann Ph.D., Michael NMN
APPLICANT: Bettler Ph.D., Bernhard NMN
APPLICANT: Jensen Ph.D., Jan E.
TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND
METHODS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
STREET: 444 So. Flower St., Suite 2000
CITY: Los Angeles
STATE: California
COUNTRY: United States
ZIP: 90071-2921
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/481,206
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/013,767
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Reiter Ph.D., Stephen E.
REGISTRATION NUMBER: 31192
REFERENCE/DOCKET NUMBER: P31 8962
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
TELEX: 9103330318
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: G1UR3
FEATURE:
NAME/KEY: CDS
LOCATION: 167..2833
US-08-481-206-5
Query Match 56.0%; Score 14; DB 1; Length 3083;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 catgtcgcaagtc 16
|||||
DB 3034 CATGTGCGCAAGTC 3021

```
RESULT 16
US-08-486-269A-5/c
; Sequence 5, Application US/08486269A
; Patent No. 5945509
; GENERAL INFORMATION:
; APPLICANT: Heinemann, Stephen F.
; APPLICANT: Boulter, James R.
; APPLICANT: Hollmann, Michael
; APPLICANT: Bettler, Bernhard
; APPLICANT: Jensen, Jan F.
; TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS
; TITLE OF INVENTION: AND METHODS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows DEMONSTRATION Version 2.00
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,269A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/013,767
; FILING DATE: 04-FEB-1993
; APPLICATION NUMBER: 07/718,575
; FILING DATE: 21-JUN-1991
; APPLICATION NUMBER: PCT/US90/06153
; FILING DATE: 25-OCT-1990
; APPLICATION NUMBER: 07/428,116
; FILING DATE: 27-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9986
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-677-1409
; TELEFAX: 619-677-1465
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3083 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; CLONE: G1UR3
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 167...2830
; OTHER INFORMATION:
;
US-08-486-269A-5

Query Match 56.0%; Score 14; DB 2; Length 3083;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 3 catgtgcagaatc 16
Db 3034 CATGTGCAAGTC 3021

RESULT 17
```

```
US-08-651-572-1/c
; Sequence 1, Application US/08651572
; Patent No. 5789228
; GENERAL INFORMATION:
; APPLICANT: Lam, D. et al.
; TITLE OF INVENTION: Endoglucanases
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESS: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/651,572
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Herron
; REGISTRATION NUMBER: 28 019
; REFERENCE/DOCKET NUMBER: 331400-48
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1662 NUCLEOTIDES
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: DNA
;
US-08-651-572-1
```

```
Query Match 52.0%; Score 13; DB 1; Length 1662;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 12 aatcacagaatt 24
Db 974 AAGTCACAGAATT 962

RESULT 18
US-09-066-544-1/c
; Sequence 1, Application US/09066544
; Patent No. 6001984
; GENERAL INFORMATION:
; APPLICANT: Lam, D. et al.
; TITLE OF INVENTION: Endoglucanases
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESS: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
```


SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/066,544
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/651,572
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Charles J. Heron
REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 331400-48
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1662 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: DNA
US-09-066-544-1

Query Match 52.0%; Score 13; DB 3; Length 1662;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aagtcacagaatt 24
|||||

DB 974 AAGTCACAGAATT 962

RESULT 19
US-08-951-086-1/c
Sequence 1, Application US/08951086
Patent No. 6074867
GENERAL INFORMATION:
APPLICANT: Lam, D. et al.
TITLE OF INVENTION: Endoglucanases
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILLILLAN,
ADDRESS: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/951,086
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/651,572
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Charles J. Heron
REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 331400-48
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1662 NUCLEOTIDES
TYPE: NUCLEIC ACID

STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: DNA
US-08-951-086-1

Query Match 52.0%; Score 13; DB 3; Length 1662;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aagtcacagaatt 24
|||||

DB 974 AAGTCACAGAATT 962

RESULT 20
US-08-765-081-4
Sequence 4, Application US/08765081
Patent No. 5798260
GENERAL INFORMATION:
APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr, J.C.
TITLE OF INVENTION: Escherichia coli 0157:H7 Epithelial Adhesin
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
STREET: 2800 Pacific First Center, 1420 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage
COMPUTER: IBM PC compatible/Pentium
OPERATING SYSTEM: MS-WINDOW 3.1
SOFTWARE: Word for Windows-6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/765,081
FILING DATE: March 26, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: June 7, 1995
APPLICATION NUMBER: US 08/265,714
FILING DATE: June 24, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Shelton, Dennis K.
REGISTRATION NUMBER: 26,997
REFERENCE/DOCKET NUMBER: CHOR-1-10286
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-206-682-8100; 1-206-224-0718 (direct)
TELEFAX: 1-206-224-0779
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2091 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
DESCRIPTION: Corresponds to SEQ ID NO:1,
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli O157: H7
STRAIN: 86-24 NAHL
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2088
US-08-765-081-4

Query Match 52.0%; Score 13; DB 1; Length 2091;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 agtcacagaatta 25
|||||
Db 692 AGTCACAGAATTA 704

RESULT 21

US-09-098-082-4

: Sequence 4, Application US/09098082
: Patent No. 6040421

: GENERAL INFORMATION:

: APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
: TITLE OF INVENTION: Escherichia coli 0157:H7 Epithelial Adhesin
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:: ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
: STREET: 2800 Pacific First Center, 1420 Fifth Avenue
: CITY: Seattle
: STATE: Washington: COUNTRY: USA
: ZIP: 98101-2347

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage

: COMPUTER: IBM PC compatible/Pentium II

: OPERATING SYSTEM: MS-Windows 95

: SOFTWARE: Word for Windows-6.0

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/09/098,082

: FILING DATE:

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: US 08/765,081

: FILING DATE: March 26, 1997

: APPLICATION NUMBER: PCT/US95/06994

: FILING DATE: June 7, 1995

: APPLICATION NUMBER: US 08/265,714

: FILING DATE: June 24, 1994

: ATTORNEY/AGENT INFORMATION:

: NAME: Shellness, Diana K.

: REGISTRATION NUMBER: 35,356

: REFERENCE/DOCKET NUMBER: CHOR-1-12402

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)

: TELEFAX: 1-206-224-0779

: INFORMATION FOR SEQ ID NO: 4:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 2091 base pairs

: TYPE: nucleic acid

: STRANDEDNESS: single

: TOPOLOGY: linear

: MOLECULE TYPE: DNA (genomic)

: DESCRIPTION: corresponds to SEQ ID NO:1,

: HYPOTHEICAL: nucleotides 3036-5126

: ANTI-SENSE: NO

: ORIGINAL SOURCE:

: ORGANISM: Escherichia coli 0157: H7

: STRAIN: 86-24 NALR

: FEATURE:

: NAME/KEY: CDS

: LOCATION: 1..2088

: US-09-098-082-4

: Query Match

: Best Local Similarity 100.0%; Score 13; DB 3; Length 2091;

: Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 agtcacagaatta 25
|||||

Db 692 AGTCACAGAATTA 704

RESULT 22

PCT-US95-06994-4
: Sequence 4, Application PC/TUS9506994

: GENERAL INFORMATION:

: APPLICANT: Children's Hospital & Medical Center

: APPLICANT: University of Washington

: APPLICANT: Washington State University Research Foundation

: APPLICANT: TARR, PHILIP I

: APPLICANT: BILGE, SIMA S

: APPLICANT: BESSER, THOMAS E

: APPLICANT: VARY JR, JAMES C

: TITLE OF INVENTION: ESCHERICHIA COLI 0157:H7 EPITHELIAL ADHESIN

: NUMBER OF SEQUENCES: 14

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC

: STREET: SUITE 2800, 1420 FIFTH AVENUE

: CITY: SEATTLE

: STATE: WASHINGTON

: COUNTRY: USA

: ZIP: WA 98101

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: Patentin Release #1.0, Version #1.30

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: PCT/US95/06994

: FILING DATE: 07-JUN-95

: CLASSIFICATION:

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: US 08/265,714

: FILING DATE: 24-JUN-1994

: ATTORNEY/AGENT INFORMATION:

: NAME: BRODERICK, THOMAS F

: REGISTRATION NUMBER: 31,332

: REFERENCE/DOCKET NUMBER: CHOR-18591

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (206) 682 8100

: TELEFAX: (206) 224 0779

: INFORMATION FOR SEQ ID NO: 4:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 2091 base pairs

: TYPE: nucleic acid

: STRANDEDNESS: single

: TOPOLOGY: linear

: MOLECULE TYPE: DNA (genomic)

: DESCRIPTION: corresponds to SEQ ID NO:1,

: HYPOTHEICAL: nucleotides 3036-5126

: ANTI-SENSE: NO

: ORIGINAL SOURCE:

: ORGANISM: Escherichia coli 0157: H7

: STRAIN: 86-24 NALR

: FEATURE:

: NAME/KEY: CDS

: LOCATION: 1..2088

: PCT-US95-06994-4

: Query Match

: Best Local Similarity 100.0%; Score 13; DB 5; Length 2091;

: Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 agtcacagaatta 25
|||||

Db 692 AGTCACAGAATTA 704

RESULT 23

US-08-484-993B-1/C

: Sequence 1, Application US/08484993B

: Patent No. 5837497

: GENERAL INFORMATION:

: APPLICANT: Harris Ph.D., Jeffrey D.

APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immuncontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,993B
FILING DATE: 09-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2214 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sus scrofa
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 12..119
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 120..2153
FEATURE:
NAME/KEY: CDS
LOCATION: 12..2153
US-08-484-993B-1

Query Match 52.0%; Score 13; DB 2; Length 2214;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacagaa 22
|||||
DB 1898 GCAAGTCACAGAA 1886

RESULT 24
US-08-484-158B-1/c
; Sequence 1, Application US/08484158B

Patent No. 5976545
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Pharmaceutical Compositions for
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,158B
FILING DATE: 07-JUNE-95
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/149,223
FILING DATE: 09-NOV-93
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-93
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-92
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 32794
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2214 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sus scrofa
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 12..119
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 120..2153
FEATURE:
NAME/KEY: CDS
LOCATION: 12..2153
US-08-484-158B-1

Query Match 52.0%; Score 13; DB 2; Length 2214;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

LOCATION: 12..2153
US-08-480-150A-1

Query Match 52.0%; Score 13; DB 2; Length 2214;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 gcaagtcacagaa 22
|||||
Db 1898 GCAAGTCACAGAA 1886

RESULT 27
US-08-458-731-1/c
Sequence 1, Application US/08458731
Patent No. 6001599
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunocontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,731
FILING DATE: 09-NOV-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2214 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sus scrofa
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: sig_peptide
NAME/KEY: mat_peptide

LOCATION: 120..2153
FEATURE:
NAME/KEY: CDS
LOCATION: 12..2153
US-08-458-731-1

Query Match 52.0%; Score 13; DB 3; Length 2214;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 gcaagtcacagaa 22
|||||
Db 1898 GCAAGTCACAGAA 1886

RESULT 28
US-08-149-223A-1/c
Sequence 1, Application US/08149223A
Patent No. 6027727
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunocontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/149,223A
FILING DATE: 09-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2214 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sus scrofa
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: sig_peptide

LOCATION: 12..119
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 120..2153
FEATURE:
NAME/KEY: CDS
LOCATION: 12..2153
US-08-149-223A-1

Query Match 52.0%; Score 13; DB 3; Length 2214;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacagaa 22
|||||
DB 1898 GCAAGTCACAGAA 1886

RESULT 29
US-08-484-993B-9/C
Sequence 9, Application US/08484993B
Patent No. 5837497
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunocontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,993B
FILING DATE: 09-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Canis familiaris
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploid
TISSUE TYPE: Ovary

CELL TYPE: Oocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 206..2353
US-08-484-993B-9

Query Match 52.0%; Score 13; DB 2; Length 2381;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacagaa 22
|||||
DB 2098 GCAAGTCACAGAA 2086

RESULT 30
US-08-484-158B-9/C
Sequence 9, Application US/08484158B
Patent No. 5976545
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Pharmaceutical Compositions for
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,158B
FILING DATE: 07-JUNE-95
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/149,223
FILING DATE: 09-NOV-93
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-93
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-92
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 32794
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Canis familiaris
DEVELOPMENTAL STAGE: Juvenile

HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 206..2353
US-08-484-158B-9

Query Match 52.0%; Score 13; DB 2; Length 2381;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacagaa 22
|||||
Db 2098 GCAAGTCACAGAA 2086

RESULT 31
US-08-484-596A-9/C
Sequence 9, Application US/08484596A
Patent No. 5981228
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunocontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,596A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/149,223
FILING DATE: 11-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Canis familiaris
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte

FEATURE:
NAME/KEY: CDS
LOCATION: 206..2353
US-08-484-596A-9

Query Match 52.0%; Score 13; DB 2; Length 2381;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacagaa 22
|||||
Db 2098 GCAAGTCACAGAA 2086

RESULT 32
US-08-480-150A-9/C
Sequence 9, Application US/08480150A
Patent No. 5989550
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunocontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,150A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 42A
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/149,223
FILING DATE: 09-NOV-1993
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Canis familiaris
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:

NAME/KEY: CDS
LOCATION: 206..2353
US-08-480-150A-9

Query Match 52.0%; Score 13; DB 2; Length 2381;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gcaagtcacagaa 22
|||||
DB 2098 GCAAGTCACAGAA 2086

RESULT 33
US-08-458-731-9/c
Sequence 9, Application US/08458731
Patent No. 6001599
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunocontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,731
FILING DATE: 09-NOV-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ. ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Canis familiaris
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: oocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 206..2353
US-08-458-731-9

Query Match 52.0%; Score 13; DB 3; Length 2381;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gcaagtcacagaa 22
|||||
DB 2098 GCAAGTCACAGAA 2086

RESULT 34
US-08-149-223A-9/c
Sequence 9, Application US/08149223A
Patent No. 6027727
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunocontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/149,223A
FILING DATE: 09-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ. ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Canis familiaris
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: oocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 206..2353
US-08-149-223A-9

Query Match 52.0%; Score 13; DB 3; Length 2381;

Best Local Similarity 100.0%; Pred. No. 30;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacagaa 22
|||||
Db 2098 GCAAGTCACAGAA 2086

RESULT 35

US-08-743-637B-35/C
; Sequence 35, Application US/08743637B
; Patent No. 5994066
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: PICARD, Francois J.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 EAST WISCONSIN AVENUE
; CITY: MILWAUKEE
; STATE: WISCONSIN
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/743,637B
; FILING DATE: 04-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,840
; FILING DATE: 11-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586,90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4500 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
US-08-743-637B-35

Query Match 52.0%; Score 13; DB 2; Length 4500;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aagtcacagaatt 24
|||||
Db 1911 AAGTCACAGAATT 1899

RESULT 36
US-08-526-840B-35/C
; Sequence 35, Application US/08526840B
; Patent No. 6001564

; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
; TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
; TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
; TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53202-4497

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,840B
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/304,732
; FILING DATE: 12-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586,90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4500 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
US-08-526-840B-35

Query Match 52.0%; Score 13; DB 3; Length 4500;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 aagtcacagaatt 24
|||||
Db 1911 AAGTCACAGAATT 1899

RESULT 37
US-08-765-081-1
; Sequence 1, Application US/08765081
; Patent No. 5798260
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli O157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44mb storage

```
COMPUTER: IBM PC compatible/Pentium
OPERATING SYSTEM: MS-Windows 3.1
SOFTWARE: Word for Windows-6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/765.081
FILING DATE: March 26, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: June 7, 1995
APPLICATION NUMBER: US 08/265,714
FILING DATE: June 24, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Shelton, Dennis K.
REGISTRATION NUMBER: 26,997
REFERENCE/DOCKET NUMBER: CHOR-1-10286
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-206-682-8100; 1-206-224-0779
TELEFAX: 1-206-224-0779
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli O157:H7
STRAIN: 86-24 NALR
IMMEDIATE SOURCE:
CLONE: PEAR
US-08-765-081-1

Query Match          52.0%; Score 13; DB 1; Length 8041;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 agtcacagaatta 25
DB 3727 AGTCACAGAATTA 3739

RESULT 38
US-09-098-082-1
Sequence 1, Application US/09098082
Patent No. 6040421
GENERAL INFORMATION:
APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
TITLE OF INVENTION: Escherichia Coli O157:H7 Epithelial Adhesin
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
STREET: 2800 Pacific First Center, 1420 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-3.5 inch, 1.44mb storage
COMPUTER: IBM PC compatible/Pentium II
OPERATING SYSTEM: MS-Windows 95
SOFTWARE: Word for Windows-6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/098.082
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/765.081
FILING DATE: March 26, 1997
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: June 7, 1995
APPLICATION NUMBER: US 08/265.714
```

```
FILING DATE: June 24, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Sheiness, Diana K.
REGISTRATION NUMBER: 35,356
REFERENCE/DOCKET NUMBER: CHOR-1-12402
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)
TELEFAX: 1-206-224-0779
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli O157:H7
STRAIN: 86-24 NALR
IMMEDIATE SOURCE:
CLONE: PEAR
US-09-098-082-1

Query Match          52.0%; Score 13; DB 3; Length 8041;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 agtcacagaatta 25
DB 3727 AGTCACAGAATTA 3739

RESULT 39
PCT-US95-06994-1
Sequence 1, Application PC/TUS9506994
GENERAL INFORMATION:
APPLICANT: Children's Hospital & Medical Center
APPLICANT: University of Washington
APPLICANT: Washington State University Research Foundation
APPLICANT: TARR, PHILIP I
APPLICANT: BILGE, SIMA S
APPLICANT: BESSER, THOMAS E
APPLICANT: VARY JR, JAMES C
TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
STREET: SUITE 2800, 1420 FIFTH AVENUE
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: WA 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: 07-JUN-95
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/265,714
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: BRODERICK, THOMAS F
REGISTRATION NUMBER: 31,332
REFERENCE/DOCKET NUMBER: CHOR-18591
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682 8100
TELEFAX: (206) 224 0779
```

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli O157:H7
STRAIN: 86-24 NALR
IMMEDIATE SOURCE:
CLONE: PEAR
PCT-US95-06994-1

Query Match 52.0%; Score 13; DB 5; Length 8041;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 agtcacagaatta 25
|||||
Db 3727 AGTCACAGATT 3739

RESULT 40
US-09-009-913-1/c
Sequence 1, Application US/09009913
Patent No. 6087485
GENERAL INFORMATION:
APPLICANT: AYS Pharmaceuticals, Inc.
TITLE OF INVENTION: Asthma Related Genes
NUMBER OF SEQUENCES: 339
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bozicevic & Reed, LLP
STREET: 285 Hamilton Ave, Suite 200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/009,913
FILING DATE: 21-JAN-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J
REGISTRATION NUMBER: 36,677
REFERENCE/DOCKET NUMBER: SEQ-4P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3231
TELEFAX: 650-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 72928 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-009-913-1

Query Match 52.0%; Score 13; DB 3; Length 72928;
Best Local Similarity 100.0%; Pred. No. 25;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 12 aagtcacagaatt 24
|||||
Db 26365 AAGTCACAGATT 26353

RESULT 41
US-09-014-065-15/c
Sequence 15, Application US/09014065
Patent No. 603854
GENERAL INFORMATION:
APPLICANT: Kurnit, David M.
APPLICANT: Chiang, Pei-Wen
APPLICANT: Wang, Chang-Ning J.
TITLE OF INVENTION: METHOD FOR DETERMINING THE COPY NUMBER OF A NUCLEIC ACID SEQUE
FILE REFERENCE: 06498/004001
CURRENT APPLICATION NUMBER: US/09/014,065
CURRENT FILING DATE: 1998-01-27
EARLIER APPLICATION NUMBER: US 08/434,474
EARLIER FILING DATE: 1995-05-04
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 15
LENGTH: 24
TYPE: DNA
ORGANISM: Homo sapiens
US-09-014-065-15

Query Match 48.0%; Score 12; DB 3; Length 24;
Best Local Similarity 100.0%; Pred. No. 14e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gcaagtcacaga 21
|||||
Db 12 GCAAGTCACAGA 1

RESULT 42
PCT-US93-03077-7
Sequence 7, Application PC/TUS9303077
GENERAL INFORMATION:
APPLICANT: Board of Regents, The University of Texas System
APPLICANT: Gaynor, Richard B.
APPLICANT: Wu, Foon Kin
TITLE OF INVENTION: PROTEIN CELLULAR FACTOR USEFUL FOR
TITLE OF INVENTION: REGULATING GENE EXPRESSION
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/03077
FILING DATE: 19930331
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/862,025
FILING DATE: April 2, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Kammerer, Patricia A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: UTFD270PCT
TELECOMMUNICATION INFORMATION:

TELEPHONE: 713-787-1540
TELEFAX: 713-749-2679
TELEX:
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 255 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: unknown
TOPOLOGY: unknown
PCN-US93-03077-7

Query Match 48.0%; Score 12; DB 5; Length 255;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 agtcacagaatt 24
|||||
Db 206 AGTCACAGAAAT 217

RESULT 43
US-08-047-041A-17/C
Sequence 17, Application US/08047041A
Patent No. 5527676
GENERAL INFORMATION:
APPLICANT: Vogelstein, Bert
APPLICANT: Baker, Suzanne J.
APPLICANT: Nigro, Janice M.
TITLE OF INVENTION: Detection of Loss of the Wild-Type p53
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, Ltd.
STREET: 1001 G Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4597
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/047,041A
FILING DATE: 22-MAR-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/928,661
FILING DATE: 17-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/446,584
FILING DATE: 06-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/330,566
FILING DATE: 29-MAR-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 01107,42917
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 279 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: N5

ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
POSITION IN GENOME:
CHROMOSOME/SEGMENT: exon 4
PUBLICATION INFORMATION:
AUTHORS: Lamb, M.Ol. Cell. Biol.
JOURNAL: Mol. Cell. Biol.
VOLUME: 6
ISSUE: 5
PAGES: 1379-1385
DATE: 1986
US-08-047-041A-17

Query Match 48.0%; Score 12; DB 1; Length 279;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcaagtcacaga 21
|||||
Db 276 GCAAGTCACAGA 265

RESULT 44
US-09-232-522-3/C
Sequence 3, Application US/09232522
Patent No. 6225117
GENERAL INFORMATION:
APPLICANT: Gateley, Maurice K.
APPLICANT: Presky, David H.
TITLE OF INVENTION: ANTIBODIES AGAINST HUMAN IL-12
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: United States
ZIP: 07110-1199
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/232,522
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Buchholz, Briana C.
REGISTRATION NUMBER: 39,123
REFERENCE/DOCKET NUMBER: CD 1048P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 973-235-6208
TELEFAX: 973-235-2363
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 308 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: mouse
CELL TYPE: Hybridoma
CELL LINE: HIL-12F3-20E11
FEATURE:
NAME/KEY: CDS
LOCATION: 1..306
US-09-232-522-3

Query Match 48.0%; Score 12; DB 4; Length 308;
 Best local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 agtcacagaatt 24
 Db 246 AGTCACAGAAATT 235

RESULT 45

US-08-047-041A-4/C
 ; Sequence 4, Application US/08047041A
 ; Patent No. 5527676
 ; GENERAL INFORMATION:
 ; APPLICANT: Vogelstein, Bert
 ; APPLICANT: Baker, Suzanne J.
 ; APPLICANT: Fearon, Eric R.
 ; APPLICANT: Nigro, Janice M.
 ; TITLE OF INVENTION: Detection of loss of the wild-type p53
 ; TITLE OF INVENTION: Gene
 ; NUMBER OF SEQUENCES: 28
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Banner & Allegretti, Ltd.
 ; STREET: 1001 G Street, N.W.
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20001 4597
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/047, 041A
 ; FILING DATE: 22-MAR-1993
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/928,661
 ; FILING DATE: 17-AUG-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/446,584
 ; FILING DATE: 06-DEC-1989
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/330,566
 ; FILING DATE: 29-MAR-1989
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Kagan, Sarah A.
 ; REGISTRATION NUMBER: 32,141
 ; REFERENCE/DOCKET NUMBER: 01107.42917
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-508-9100
 ; TELEFAX: 202-508-9299
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 311 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; POSITION IN GENOME:
 ; CHROMOSOME/SEGMENT: exon 4
 ; PUBLICATION INFORMATION:
 ; AUTHORS: Buchman, V. L.
 ; TITLE: A variation in the structure of the
 ; TITLE: protein-coding region of the human p53 gene
 ; JOURNAL: Gene

VOLUME: 70
 ; PAGES: 245-252
 ; DATE: 1988
 ; US-08-047-041A-4

Query Match 48.0%; Score 12; DB 1; Length 311;
 Best local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gcaagtcacaga 21
 Db 293 GCAAGTCACAGA 282

Search completed: October 9, 2001, 15:55:17
 Job time: 13193 sec

Wed Oct 10 07:45:32 2001

us-09-396-196f-1.oli.rni

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:27:58 ; Search time 5323.87 Seconds
(without alignments)
44.389 Million cell updates/sec

Title: US-09-396-196f-1

Perfect score: 25

Sequence: 1 gacatgtgcgaagtcacagaatta 25

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 10228115 segs, 4726426750 residues

Word size : 9

Total number of hits satisfying chosen parameters: 467815

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est148:*
 118: gb_est149:*
 119: gb_est150:*
 120: gb_est151:*
 121: gb_est152:*
 122: gb_est153:*
 123: gb_est154:*
 124: gb_est155:*
 125: gb_est156:*
 126: gb_est157:*
 127: gb_est158:*
 128: gb_est159:*
 129: gb_est160:*
 130: gb_est161:*
 131: gb_est162:*
 132: gb_est163:*
 133: gb_est164:*
 134: gb_est165:*
 135: gb_est166:*
 136: gb_est167:*
 137: gb_est168:*
 138: gb_est169:*
 139: gb_est170:*
 140: gb_est171:*
 141: gb_est172:*
 142: gb_est173:*
 143: gb_est174:*
 144: gb_est175:*
 145: gb_est176:*
 146: gb_est177:*
 147: gb_est178:*
 148: gb_est179:*
 149: gb_est180:*
 150: gb_est181:*
 151: gb_est182:*
 152: gb_est183:*
 153: gb_est184:*
 154: gb_est185:*
 155: gb_est186:*
 156: gb_est187:*
 157: gb_est188:*
 158: gb_est189:*
 159: gb_est190:*
 160: gb_est191:*
 161: gb_est192:*
 162: gb_est193:*
 163: gb_est194:*
 164: gb_est195:*
 165: gb_est196:*
 166: gb_est197:*
 167: gb_est198:*
 168: gb_est199:*
 169: gb_est200:*
 170: gb_est201:*
 171: gb_est202:*
 172: gb_est203:*
 173: gb_est204:*
 174: gb_est205:*
 175: gb_est206:*
 176: gb_est207:*
 177: gb_est208:*
 178: gb_est209:*
 179: gb_est210:*
 180: gb_est211:*
 181: gb_est212:*
 182: gb_est213:*
 183: gb_est214:*
 184: gb_est215:*
 185: gb_est216:*
 186: gb_est217:*
 187: gb_est218:*
 188: gb_est219:*
 189: gb_est220:*

190: gb_est110:*
 191: gb_est111:*
 192: gb_est112:*
 193: gb_est113:*
 194: gb_est114:*
 195: gb_est115:*
 196: gb_est116:*
 197: gb_est117:*
 198: gb_est118:*
 199: gb_est119:*
 200: gb_est120:*
 201: gb_est121:*
 202: gb_est122:*
 203: gb_est123:*
 204: gb_est124:*
 205: gb_est125:*
 206: gb_est126:*
 207: gb_est127:*
 208: gb_est128:*
 209: gb_est129:*
 210: gb_est130:*
 211: gb_est131:*
 212: gb_est132:*
 213: gb_est133:*
 214: gb_est134:*
 215: gb_est135:*
 216: gb_est136:*
 217: gb_est137:*
 218: gb_est138:*
 219: gb_est139:*
 220: gb_est140:*
 221: gb_est141:*
 222: gb_est142:*
 223: gb_est143:*
 224: gb_est144:*
 225: gb_est145:*
 226: gb_est146:*
 227: gb_est147:*
 228: gb_est148:*
 229: gb_est149:*
 230: gb_est150:*
 231: gb_est151:*
 232: gb_est152:*
 233: gb_est153:*
 234: gb_est154:*
 235: gb_est155:*
 236: gb_est156:*
 237: gb_est157:*
 238: gb_est158:*
 239: gb_est159:*
 240: gb_est160:*
 241: gb_est161:*
 242: gb_est162:*
 243: gb_est163:*
 244: gb_est164:*
 245: gb_est165:*
 246: gb_est166:*
 247: gb_est167:*
 248: gb_est168:*
 249: gb_est169:*
 250: gb_est170:*
 251: gb_est171:*
 252: gb_est172:*
 253: gb_est173:*
 254: gb_est174:*
 255: gb_est175:*
 256: gb_est176:*
 257: gb_est177:*
 258: gb_est178:*

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	17	68.0	357	140	BE800706	BE800706 sq97c06.y
2	17	68.0	388	140	AW781150	AW781150 s189c06.y
3	17	68.0	751	138	BE58717	BE58717 GM700007A
4	16	64.0	299	138	BB450863	BB450863 BB450863
5	16	64.0	418	150	BF556014	BF556014 UI-R-A1-d
6	15	60.0	214	121	AW871559	AW871559 ra79a04.y
7	15	60.0	233	139	BE763893	BE763893 RC4-NT005
8	15	60.0	253	139	BE763892	BE763892 RC4-NT005
9	15	60.0	256	5	AA317110	AA317110 EST19001
10	15	60.0	261	191	Z45844	Z45844 HSC2VE041 n
11	15	60.0	263	126	BB141316	BB141316 BB141316
12	15	60.0	263	159	BF745516	BF745516 CM2-BT082
13	15	60.0	282	159	BF800965	BF800965 QV4-C1010
14	15	60.0	286	6	AA344205	AA344205 EST50086
15	15	60.0	286	256	H53227	B53227 C1T-HSP-200
16	15	60.0	317	159	BF743951	BF743951 CM2-BT082
17	15	60.0	327	188	T30901	T30901 EST24536 Hu
18	15	60.0	335	159	BF743964	BF743964 CM2-BT082
19	15	60.0	346	159	BF743954	BF743954 CM2-BT082
20	15	60.0	349	154	BG466199	BG466199 ESTE825
21	15	60.0	353	10	AA648449	AA648449 ns22f10.s
22	15	60.0	370	188	T07051	T07051 EST04940 Fe
23	15	60.0	402	23	A1694429	A1694429 wd83f06.x
24	15	60.0	404	17	A1203847	A1203847 gf75h03.x
25	15	60.0	405	168	BF726110	BF726110 by01c01.x
26	15	60.0	406	153	BG413230	BG413230 ESTE818
27	15	60.0	416	104	A1942464	A1942464 wp08c05.x
28	15	60.0	420	3	AA172188	AA172188 zp29f06.s
29	15	60.0	427	224	AO142435	AO142435 HS-3064-A
30	15	60.0	429	121	AM870882	AM870882 ra51h09.y
31	15	60.0	454	224	AO086771	AO086771 HS-2175-B
32	15	60.0	455	14	AA973493	AA973493 o045e05.s
33	15	60.0	456	24	A1756736	A1756736 ESTE824
34	15	60.0	457	19	A1361386	A1361386 qy47f01.x
35	15	60.0	464	155	BG562371	BG562371 ESTE805
36	15	60.0	471	22	A1635259	A1635259 fz79h07.x
37	15	60.0	485	172	BG020076	BG020076 dc47c07.x
38	15	60.0	487	154	BG516037	BG516037 ESTE848
39	15	60.0	498	169	BF759335	BF759335 286217 MA
40	15	60.0	498	243	AZ398330	AZ398330 1M0163N15
41	15	60.0	502	12	AA846453	AA846453 aj36b03.s
42	15	60.0	502	23	A1672463	A1672463 wa03e05.x
43	15	60.0	503	22	A1634722	A1634722 fz09e04.x
44	15	60.0	512	115	AM418528	AM418528 xc32c07.x
45	15	60.0	512	233	AO801291	AO801291 HS-2088-A

ALIGNMENTS

RESULT 1
BE800706
DEFINITION BE800706 357 bp mRNA
sq97c06.y1 Gm-c1049 glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1049-1091 5' similar to TR:082150 082150 CHLOROPLAST FTSH
ACCESSION BE800706
VERSION BE800706.1
KEYWORDS GI:10231818
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Corryell,V., Khanna
AUTHORS A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,

TITLE
JOURNAL
COMMENT
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or
IntroGenomesystems.com web site: www.genomesystems.com
Putative full length read
vector to vector length is 364
High quality sequence stop: 340.

FEATURES

source
location/Qualifiers
1..357
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1049-1091"
/clone_1lb="Gm-c1049"
/tissue_type="Whole seedlings of greenhouse grown plants"
/dev_stage="3 week old"
/lab_host="DH10B"
/note="Vector: pluescript II SK+; Site_1: EcoRI; Site_2: XhoI; The Clark NTL was constructed and seed was provided by Dr. J. Specht, University of Nebraska (Shoemaker and Specht, 1995). The cDNA library was constructed from mRNA isolated from whole seedlings of 3 week old greenhouse grown plants. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site and a 3' anchor. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). The library was constructed in cooperation with Dr. Paul Keim's laboratory at Northern Arizona University."

BASE COUNT

169 a 53 c 47 g 88 t

Query Match 68.0%; Score 17; DB 140; Length 357;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 ccgaatcacagaatt 25
DB 138 CGCAAGTCACAGATT 154
|||||

RESULT 2
AW781150
LOCUS AW781150 388 bp mRNA
DEFINITION AW781150 s189c06.y1 Gm-c1037 glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1037-1235 5' similar to TR:082150 082150 CHLOROPLAST FTSH
ACCESSION AW781150
VERSION AW781150.1
KEYWORDS GI:7795753
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE I (bases 1 to 388)

AUTHORS
Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corvelli, V., Khanna, A., Bolla, B., Marita, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ralster, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.
Public Soybean EST Project
Unpublished (1999)

TITLE
JOURNAL
COMMENT

FEATURES
source
Location/Qualifiers
1. .388
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl037-1235"
/clone_lib="Gm-cl037"
/tissue_type="fully expanded leaves of greenhouse grown plants"
/dev_stage="2 week old"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI. This cDNA library was constructed from mRNA isolated from fully expanded leaves of greenhouse grown plants that were 2 weeks old. The library was prepared using the life technologies psupertscript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly(dt) sequence with a NotI restriction site. SalI linkers adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli Electro-Max DH10B host cells. This library was constructed in the laboratory of Dr. Lila Vodkin by Anu Khanna at the University of Illinois at Urbana-Champaign. email: l-vodkin@uiuc.edu"

BASE COUNT
123 a 67 c 92 g 106 t

ORIGIN

Query Match 68.0%; Score 17; DB 120; Length 388;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 cgcagtcacagaatta 25
|||||

DB 334 cgcagtcacagaatta 350

RESULT 3
BE658717 751 bp mRNA EST 06-SEP-2000
LOCUS GM700007A10F9 Gm-r1070 Glycine max cDNA clone Gm-r1070-2561 3',
DEFINITION mRNA sequence.
ACCESSION BE658717
VERSION BE658717.1 GI:9984609
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 751)

AUTHORS
Vodkin, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Corvelli, V., Erpelting, J., Rapp, C., Shoop, E., Pardini, J., Liu, L., and Lewin, H.
A Functional Genomics Program for Soybean (NSF 9872565)
Unpublished (1999)
Other ESTs: AW156684 corresponding to Gm-cl015-2632 (5')
Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu

TITLE
JOURNAL
COMMENT

FEATURES
source
Location/Qualifiers
1. .751
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="Gm-r1070-2561"
/clone_lib="Gm-r1070"
/note="The library Gm-r1070 is a sequence-driven, rerecked set of 9,216 clones selected from cDNA libraries from various tissues and stages of development of soybean that represent 2,639 sequences from immature cotyledons, 1,770 from immature seed coats, 3,938 from flowers, and 865 from young pods. The 5' ESTs of the source clones from the different libraries was used to select singletons, or a representative of each contig, which were rerecked to form library Gm-r1070. The cDNA clones of the rerecked Gm-r1070 library were then sequenced at the 3' end. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, http://www.cbc.umn.edu/ResearchProjects/soybean/index.html. Rerecking was performed by Genome Systems, St. Louis, http://www.genomesystems.com, and 3' sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, http://www.life.uiuc.edu/biotech/keck.html. Note: The corresponding 5' EST from each clone in the Gm-r1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'."

BASE COUNT
193 a 171 c 131 g 226 t 30 others

ORIGIN

Query Match 68.0%; Score 17; DB 138; Length 751;
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 cgcagtcacagaatta 25
|||||

DB 112 cgcagtcacagaatta 96

RESULT 4
BA450863 299 bp mRNA EST 21-JUL-2000
LOCUS BA450863 RIKEN full-length enriched, 12 days embryo spinal ganglion
DEFINITION Mus musculus cDNA clone D130006117 3', mRNA sequence.
ACCESSION BA450863
VERSION BA450863.1 GI:9311898
KEYWORDS EST.
SOURCE house mouse.

Query Match 64.0%; Score 16; DB 150; Length 418;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacatgtcgcagacta 16
 |||
 Db 110 GACATTGTCCGACAGTC 95

RESULT 6
 AM871559/c 214 bp mRNA EST 05-APR-2001
 LOCUS ta9604.y1 Bird-Rao Meloidogyne incognita J2 Meloidogyne incognita
 DEFINITION CDNA 5' similar to WP:C26D10.4 CE03027 ; mRNA sequence.
 AM871559
 VERSION AM871559.1 GI:8005612
 KEYWORDS EST.
 SOURCE southern root-knot nematode.
 ORGANISM Meloidogyne incognita
 Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
 Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

REFERENCE 1 (bases 1 to 214)
 AUTHORS McCarter,J., Clifton,S., Chiapelli,B., Page,D., Martin,J., Wylie,T.,
 Dante,M., Marra,M., Hillier,L., Kucaba,T., Treising,B., Bowers,T.,
 Gibbons,M., Ritey,E., Bennett,D., Franklin,C., Tsagarisvilli,R.,
 Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe,
 M., Allen,M., Person,B., Swaller,T., Harvey,N., Schuck,R., Kohn,S.,
 Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
 Wilson,R.
 The Washington Univ. Nematode EST Project, 1999
 Unpublished (1999)
 JOURNAL Contact: McCarter JP
 COMMENT The Washington Univ. Nematode EST Project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

FEATURES
 source
 1..214
 /organism="Meloidogyne incognita"
 /db_xref="taxon:6306"
 /clone_lib="Bird-Rao Meloidogyne incognita J2"
 /dev_stage="enriched for 2nd stage juveniles"
 /lab_host="XLR01"
 /note="Vector: ZAP express - PBKCMV (Stratagene); Site_1:
 EcoRI; Site_2: XhoI; Oligo (dT) primed library. cDNA was
 constructed and cloned unidirectionally into the vector
 within the 5' EcoRI and 3' XhoI sites. This library was
 constructed by Dr. Uma Rao and Dr. David Bird at North
 Carolina state University."
 BASE COUNT 59 a 52 c 19 g 84 t
 ORIGIN

Query Match 60.0%; Score 15; DB 121; Length 214;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaatta 25
 |||
 Db 184 CAAGTCACAGAAATTA 170

RESULT 7
 BE763893 233 bp mRNA EST 19-SEP-2000
 LOCUS RC4-NT0054-120600-015-g05 NT0054 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BE763893
 ACCESSION BE763893.1 GI:10193817
 VERSION
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 233)
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,M. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 JOURNAL 20202663
 MEDLINE
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?rl=et2-RC4-NT0054-120600-015-g05&rl3-2000-06-12&rl4=1>)
 Seq primer: puc 18 forward
 High quality sequence stop: 233.

FEATURES
 source
 1..233
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NT0054"
 /dev_stage="Adult"
 /note="Organ: nervous-tumor; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESPES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 88 a 27 c 52 g 66 t
 ORIGIN

Query Match 60.0%; Score 15; DB 139; Length 233;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaatta 25
 |||
 Db 94 CAAGTCACAGAAATTA 108

RESULT 8
 BE763892 253 bp mRNA EST 19-SEP-2000
 LOCUS RC4-NT0054-120600-015-g03 NT0054 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BE763892
 ACCESSION BE763892.1 GI:10193816
 VERSION
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 253)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.P., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 JOURNAL MEDLINE COMMENT
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-RC4-WT0054-120
 600-015-903&t3=2000-06-12&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 12
 High quality sequence stop: 253.
 Location/Qualifiers
 1..253
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="WT0054"
 /dev_stage="Adult"
 /note="Organ: nervous tumor; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 95 a 31 c 56 g 71 t
 ORIGIN

Query Match 60.0%; Score 15; DB 139; Length 253;
 Best Local Similarity 100.0%; Pred. No. 82;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caagtcacagaatta 25
 ||||||||||||
 Db 114 CAAGTCACAGAAATTA 128

RESULT 9
 LOCUS AA317110 256 bp mRNA EST 19-APR-1997
 DEFINITION EST19001 Lung Homo sapiens cDNA 5' end, mRNA sequence.
 ACCESSION AA317110
 VERSION AA317110.1 GI:11969448
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 256)
 Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fulmer, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Giordex, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M., Kelley, J.C., Liu, L.-T., Marmaros, S.M., Merrick, J.M., Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligriano, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,

TITLE
 JOURNAL MEDLINE COMMENT
 Other ESTs: THG99418
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (http://www.tigr.org/tdb/hgi/hgi.html)
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..256
 /organism="Homo sapiens"
 /db_xref="ATCC (lnhost):117571"
 /db_xref="taxon:9606"
 /clone_lib="Lung"
 /dev_stage="adult"
 /note="Organ: Lung; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"
 BASE COUNT 73 a 57 c 48 g 75 t 3 others
 ORIGIN

Query Match 60.0%; Score 15; DB 5; Length 256;
 Best Local Similarity 100.0%; Pred. No. 82;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caagtcacagaatta 25
 ||||||||||||
 Db 189 CAAGTCACAGAAATTA 203

RESULT 10
 LOCUS 245844/c 261 bp mRNA EST 14-NOV-1994
 DEFINITION HSCZVE041 normalized infant brain cDNA Homo sapiens cDNA clone c-zve041, mRNA sequence.
 ACCESSION 245844
 VERSION 245844.1 GI:575078
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 261)
 Aufray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Devignes, M.D., Duprat, S., Houllatte, R., Jumeau, M.N., Lamy, B., Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pietu, G., Pouliot, Y., Sebastiani, K., Kachis, C. and Tessier, A.
 IMAG: molecular integration of the analysis of the human genome and its expression
 C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
 95277534
 JOURNAL MEDLINE COMMENT
 Contact: Genethon
 Genexpress-Genethon
 Genethon Centre de recherche sur le Genome Humain
 1, rue de l'Internationale, Bp60 91002 Evry Cedex, FRANCE
 Tel: 33169472800
 Fax: 33160778698

Email: genexpress@genethon.fr
Single read.
Genexpress_library_id: C; Genexpress_sequence_id: ylc-zve04
Seq primer: (-21)M3_universasl.

FEATURES

source

1. 263
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-zve04"
/clone_lib="normalized infant brain cDNA"
/sex="female"
/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lafmid BA; Site: HindIII;
Site 2: NotI; sex=Female; dev_stage=3 months old;
isolate=muscular atrophy patient; tissue=total brain
total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lafmid BA vector. Clone library from B.Soures, Psychiatry
Dept. Columbia University, USA. Normalization method:
Bento Soares, P.N.A.S in press"

BASE COUNT

77 a 56 c 31 g 97 t

ORIGIN

Query Match 60.0%; Score 15; DB 191; Length 261;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caagtcacagaatta 25
|||||

Db 138 CAAGTCACAGAATTA 124

RESULT 11

LOCUS B141316 263 bp mRNA EST 28-JUN-2000
DEFINITION B141316 RIKEN full-length enriched, adult female vagina Mus
musculus cDNA clone 993007004 3', mRNA sequence.

ACCESSION B141316 GI:8796253

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE
JOURNAL
COMMENT

Mus musculus
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 263)
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Ono, T., Owa, C., Saito, H., Sakai, C., Satoh, K., Shibata, K., Shibata,
Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suganara, I.,
Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya,
T., Tsunoda, Y., Watanishi, A., Watanabe, S., Yamamura, T., Yamataka, I.,
Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino,
M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Konno, H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rct.riken.go.jp/
URL: http://genome.rct.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki,
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

FEATURES

source

Thermostabilization and thermoactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Katsunari, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (<http://genome.rct.riken.go.jp>) for
further details.

BASE COUNT 69 a 56 c 43 g 95 t

ORIGIN

Query Match 60.0%; Score 15; DB 126; Length 263;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caagtcacagaatta 25
|||||

Db 62 CAAGTCACAGAATTA 48

RESULT 12

LOCUS BF745516 263 bp mRNA EST 10-JAN-2001
DEFINITION BF745516 C2-BF0826-021100-465-b05 BF0826 Homo sapiens cDNA, mRNA sequence.

ACCESSION BF745516 GI:12072192

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 263)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed

JOURNAL
MEDLINE
COMMENT

sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

FEATURES

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=CM2&t2=CM2-BF0826-
021100-469-h05&t3=2000-11-02&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 139
High quality sequence stop: 263.
Location/Qualifiers
1. 263

BASE COUNT

77 a 55 c 60 g 71 t
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BF0826"
/dev_stage="Adult"
/note="Organ: Breast; Vector: puc18; Site_1: Sma1; Site_2:
Sma1; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

Query Match 60.0%; Score 15; DB 169; Length 263;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 10 gcaagtcacagaatt 24
|||||
Db 217 GCAAGTCACAGATT 231

RESULT 13
BF800965/c 282 bp mRNA EST 12-JAN-2001
LOCUS
DEFINITION QV4-C10100-231000-457-d12 C10100 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF800965
VERSION BF800965.1 GI:12129954
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 282)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Bionesi,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.,J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

TITLE
JOURNAL
MEDLINE
COMMENT

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

FEATURES

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=QV4&t2=QV4-C10100-
231000-457-d12&t3=2000-10-23&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 59
High quality sequence stop: 282.
Location/Qualifiers
1. 282

BASE COUNT

68 a 75 c 67 g 72 t
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="C10100"
/dev_stage="Adult"
/note="Organ: colon; ins: Vector: puc18; Site_1: Sma1;
Site_2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

Query Match 60.0%; Score 15; DB 169; Length 282;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 4 atgttcgaagtcac 18
|||||
Db 51 ATTGTGCAAGTCAC 37

RESULT 14
AA344205 286 bp mRNA EST 21-APR-1997
LOCUS
DEFINITION EST50086 gall bladder I Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION AA344205
VERSION AA344205.1 GI:1996443
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 286)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
O., Lee,N.R., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
C.J., Lee,N.R., Blake,J.A., Brandon,R.C., Man-Wei,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
L.M., Fitzhugh,W.M., Fitchman,J.L., Geoghegan,N.S., Glodek,A.,
Gnethm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,
Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,
Small,R.V., Spriggs,T.A., Uterback,T.R., Weidman,J.F., Li,Y.,
Bednarek,K.D., P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,M.W.,
Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
Weli,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon
M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
Venter,J.C.

TITLE
JOURNAL
MEDLINE
COMMENT

Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280
Other-ESTs: THC99418
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056

Fax: 3018699423
Email: arkerlavet@tigr.org

For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.htm>)
Seq primer: M13 Reverse.

FEATURES

source

Location/Qualifiers

1..286
/organism="Homo sapiens"
/db_xref="AYC (Inhost):145738"
/db_xref="taxon:9606"
/clone_lib="Gall bladder 1"
/sex="female"
/dev_stage="adult, 25 yrs"
/note="Organ: gall bladder; Vector: pBluescript SK-
Site_1: EcoRI; Site_2: XhoI"
Site_1: EcoRI; Site_2: XhoI

BASE COUNT 83 a 62 c 48 g 91 t 2 others
ORIGIN

Query Match 60.0%; Score 15; DB 6; Length 286;
Best Local Similarity 100.0%; Pred. No. 83;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caagtcacagaatta 25
|||||
Db 197 CAAGTCACAGAAATTA 211

RESULT 15

LOCUS B53227

DEFINITION CIT-HSP-2009H2.1R CIT-HSP Homo sapiens genomic clone 2009H2. DNA sequence.

ACCESSION B53227

VERSION B53227.1 GI:2607561

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
'K', Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building
Unpublished (1997)

JOURNAL Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES

source

Location/Qualifiers

1..292
/organism="Homo sapiens"
/db_xref="GDB:7040812"
/db_xref="taxon:9606"
/clone_lib="2009H2"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
HindIII"

BASE COUNT 58 a 64 c 90 g 80 t

ORIGIN

Query Match 60.0%; Score 15; DB 169; Length 317;
Best Local Similarity 100.0%; Pred. No. 83;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gcaagtcacagaatt 24
|||||
Db 45 GCAAGTCACAGAAATT 31

Query Match 60.0%; Score 15; DB 256; Length 292;
Best Local Similarity 100.0%; Pred. No. 83;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caagtcacagaatta 25
|||||
Db 248 CAAGTCACAGAAATTA 234

RESULT 16

LOCUS BF743951

DEFINITION CM2-BT0826-181000-428-c02 BT0826 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF743951
VERSION BF743951.1 GI:12070627
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 317)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Negai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL MEDLINE

COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?l1-CM2&t2-CM2-BT0826-181000-428-c02&t3-2000-10-18&t4=1>)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 317.

FEATURES

source

Location/Qualifiers

1..317
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0826"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 96 a 71 c 49 g 101 t

ORIGIN

RESULT 17
T30901/c 327 bp mRNA EST 06-SEP-1995
LOCUS EST62536 Human Brain Homo sapiens cDNA 5' end similar to None, mRNA
DEFINITION sequence.
ACCESSION T30901
VERSION T30901.1 GI:612999
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 327)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W., Clayton,R.A.,
Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
L.M., Fitchugh,W.M., Fritchman,J.L., Geoghegan,N.S.M., Glodek,A.,
Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkleut,P.S., Kelley,J.M.,
Klimak,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palanges,R.F., McDonald,L.A., Nguyen,D.T., Pellegrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Uterback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.-J.,
Dimke,D., Peng,P., Ferrite,A., Fischer,C., Hastings,G.A., He,W.-W.,
Hu,J.-S., Greene,J.M., Gruber,J., Hudson,P., Kim,A., Kozak,D.L.,
Kunsch,C., Ji,H., Li,H., Melsner,P.S., Olsen,H., Raymond,L., Wei
Y.-F., Wang,J., Xu,C., Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon
M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
Venter,J.C.
TITLE Initial Assessment of Human Gene Diversity and Expression Patterns
JOURNAL Based Upon 83 Million Basepairs of cDNA Sequence
MEDLINE Nature 377, 3-174 (1995)
COMMENT Other ESTs: THC10632
Contact: Venter, J.C.
The Institute for Genomic Research
932 Clopper Rd, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: tdbinfo@tdb.tigr.org
For clone availability, additional sequence and expression
information related to this EST, please contact the TIGR Database
(tdbinfo@tdb.tigr.org)
Seq primer: M13 Reverse.
FEATURES
source Location/Qualifiers
1..327
/organism="Homo sapiens"
/db_xref="ATCC (inhost):100882"
/db_xref="taxon:9606"
/clone_lib="Human Brain"
/note="Organ: brain"
BASE COUNT 89 a 75 c 47 g 115 t 1 others
ORIGIN

Query Match 60.0%; Score 15; DB 188; Length 327;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 11 caagtcacagaatt 25
|||||
Db 295 CAAAGTCACAGAATT 281

RESULT 18
BF743964/c 335 bp mRNA EST 10-JAN-2001
LOCUS BF743964
DEFINITION CM2-BT0826-181000-428-h04 BT0826 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF743964
VERSION BF743964.1 GI:12070640
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 335)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Britones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the RAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=CM2&t2=CM2-BT0826-
181000-428-h04&t3=2000-10-18&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 40
High quality sequence stop: 335.
FEATURES
source Location/Qualifiers
1..335
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0826"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site:1; Smal; Site:2;
Smal: A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the puc18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions.
BASE COUNT 101 a 76 c 55 g 103 t
ORIGIN

Query Match 60.0%; Score 15; DB 169; Length 335;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 10 gcaagtcacagaatt 24
|||||
Db 61 CCAAGTCACAGAATT 47

RESULT 19
BF743954/c 346 bp mRNA EST 10-JAN-2001
LOCUS BF743954
DEFINITION CM2-BT0826-181000-428-d07 BT0826 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF743954
VERSION BF743954.1 GI:12070630
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 346)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Britones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

TITLE Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2&tl2=CM2-BF0826-
 181000-428-d07&tl3=2000-10-18&tl4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 43
 High quality sequence stop: 346.

FEATURES

source

1..346
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="BF0826"
 /dev_stage="Adult"
 /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
 SmaI; A mini-library was made by cloning products derived
 from ORESTES PCR (U.S. Letters Patent application No. 196
 ,716 - Ludwig Institute for Cancer Research) profiles
 into the pUC 18 vector. Reverse transcription of tissue
 mRNA and cDNA amplification were performed under low
 stringency conditions."

BASE COUNT 103 a 78 c 54 g 111 t
 ORIGIN

Query Match 60.0%; Score 15; DB 169; Length 346;
 Best Local Similarity 100.0%; Pred. No. 84;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gcaagtcacagaatt 24
 ||||||||||||||||
 Db 73 GCAAGTCACAGAATT 59

RESULT 20
 BG466199/c 349 bp mRNA EST 20-MAR-2001
 LOCUS E1ESTred25c05.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
 DEFINITION tenella cDNA 5' similar to TR:088349 088349 LATENT TGF BETA BINDING
 PROTEIN. ; mRNA sequence.

ACCESSION BG466199.1 GI:13395174
 VERSION BG466199
 KEYWORDS EST.
 SOURCE Eimeria tenella.
 ORGANISM Eimeria tenella.
 Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
 Eimeria.

REFERENCE 1 (bases 1 to 349)
 AUTHORS Liberators,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
 Martin,J., Wylie,T., Underwood,K., Steptoe,W., Theising,B., Allen
 M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
 N., Schurk,R., Rilter,E., Kolm,S., Florence,N., Shin,T., Jackson
 Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
 Washu-Merck Eimeria tenella project
 Unpublished (1999)
 CONTACT: David Sibley, Ph.D.
 Washu-Merck Eimeria tenella project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

COMMENT

Email: est@watson.wustl.edu
 Contact David Sibley (toxoest@borcim.wustl.edu) for further
 information relating to organism, libraries, or clone availability.
 Seq primer: -40BP from Gibco
 High quality sequence stop: 318.
 Location/Qualifiers

FEATURES

source

1..349
 /organism="Eimeria tenella"
 /strain="1518"
 /db_xref="taxon:5802"
 /clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
 /dev_stage="Sporozoite stage"
 /lab_host="SOLR"
 /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
 ; Sporozoites were obtained from in vitro sporulated and
 excysed oocysts of E. tenella grown in chickens. cDNA was
 synthesized from poly mRNA using an oligo-dT primer
 containing a XhoI site. Following second strand synthesis,
 EcoRI adaptors were ligated to the cDNA and products were
 size-selected on sephacryl S500. The cDNA were ligated to
 EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were
 converted to phagemids by mass excision using EXASist
 helper phage and E.coli SOLR cell (Stratagene). Clones
 were selected by negative hybridization against a pool of
 overrepresented ESTs (N=10, from 1682 previous reads).
 Insert sizes range from 1.2-2.9Kb. The library may contain
 a small percentage of host or bacterial contaminants."

BASE COUNT 75 a 107 c 103 g 64 t
 ORIGIN

Query Match 60.0%; Score 15; DB 154; Length 349;
 Best Local Similarity 100.0%; Pred. No. 84;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gacatttcgcaagt 15
 ||||||||||||||||
 Db 328 GACATTTCGCAAGT 314

RESULT 21
 AA648449 353 bp mRNA EST 13-NOV-1997
 LOCUS ns22f10.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1184395 3'
 DEFINITION similar to contains element LTR5 repetitive element ; mRNA
 sequence.

ACCESSION AA648449
 VERSION AA648449
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 353)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Unknown library type
 Insert length: 2027 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 334.
 Location/Qualifiers

FEATURES

source

1..353
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1184395"
 /clone_lib="NCI_CGAP_GCB1"
 /tissue_type="germinal center B cell"
 /lab_host="DH10B"
 /note="Vector: p77r3D-Pac (Pharmacia) with a modified

polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer
 15'-GTGACCATCTGACAGTGGAGCGCCGCCATTTTCTTTTCTTTT-3'
 1. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT
 ORIGIN

131 a 49 c 39 g 134 t

Query Match 60.0%; Score 15; DB 10; Length 353;
 Best Local Similarity 100.0%; Pred. No. 84;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaatta 25
 |||||||
 Db 249 CAAGTCACAGAATTA 263

RESULT 22
 LOCUS T07051 370 bp mRNA EST 30-JUN-1993
 DEFINITION EST04940 Fetal brain, Strata gene (cat936206) Homo sapiens cDNA
 clone HBBEC79, mRNA sequence.
 ACCESSION T07051
 VERSION T07051.1 GI:318200
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 370)
 AUTHORS Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
 TITLE 3,400 expressed sequence tags identify diversity of transcripts from human brain
 JOURNAL Nature Genet. 4, 256-267 (1993)

MEDLINE 93364420
 COMMENT Contact: Adams, MD
 The Institute for Genomic Research
 932 Clopper Road, Gaithersburg, MD 20878
 Tel: 3018699056
 Fax: 3018699423
 Email: mdadams@tigr.org
 Seq primer: M13-21.

FEATURES
 source Location/Qualifiers

1..370

/organism="Homo sapiens"
 /db_xref="ATCC (lnhost):83698"
 /db_xref="taxon:9606"

/clone="HBBEC79"

/clone_lib="Fetal brain, Strata gene (cat936206)"
 /note="Vector: LambdaZAP-II; 17-18 wk gestation, female;
 oligo-dT + random primed cDNA synthesis; lambdaZAP-II
 vector, 1.0kb average insert size."

BASE COUNT 137 a 58 c 68 g 106 t 1 others
 ORIGIN

Query Match 60.0%; Score 15; DB 188; Length 370;
 Best Local Similarity 100.0%; Pred. No. 84;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaatta 25
 |||||||
 Db 113 CAAGTCACAGAATTA 127

RESULT 23
 LOCUS A1694429 402 bp mRNA EST 02-JUN-1999
 DEFINITION w83106.X1 NCI-CGAP-Lu24 Homo sapiens cDNA clone IMAGE:2338211 3', mRNA sequence.
 ACCESSION A1694429
 VERSION A1694429.1 GI:4971769
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 402)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILND at:
 www.bio.lnl.gov/bbrp/image/image.html
 Seq primer: -400P from GIBCO.

FEATURES
 source Location/Qualifiers

1..402

/organism="Homo sapiens"
 /db_xref="taxon:9606"

/clone="IMAGE:2338211"

/clone_lib="NCI-CGAP_Lu24"

/tissue_type="Carcinoid"

/lab_host="DH10B"

/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI-CGAP Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones IDs 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 160 a 60 c 69 g 113 t
 ORIGIN

Query Match 60.0%; Score 15; DB 23; Length 402;
 Best Local Similarity 100.0%; Pred. No. 84;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaatta 25
 |||||||
 Db 342 CAAGTCACAGAATTA 356

RESULT 24
 LOCUS A1203847 404 bp mRNA EST 27-JAN-1999
 DEFINITION qf75h03.x1 Soares_testlis_NHT Homo sapiens cDNA clone IMAGE:1755893
 3', mRNA sequence.
 ACCESSION A1203847
 VERSION A1203847.1 GI:3756453
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 404)
 REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL
COMMENT
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 563 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 394.
Location/Qualifiers

FEATURES

source

1. 404
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1755893"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGACAGTGGAGCGGCGCCCAATTGTTTTTTTTTTT 3']
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cos5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 161 a 61 c 69 g 113 t
ORIGIN

Query Match 60.0%; Score 15; DB 17; Length 404;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 caagtcacagaatta 25
|||||
Db 342 CAAGTCACAGAATTA 356

RESULT 25
BF726110 405 bp mRNA EST 05-JAN-2001
LOCUS by01c01.x1 Human lens cDNA (Un-normalized, unamplified): BY Homo
DEFINITION sapiens cDNA clone by01c01 3', mRNA sequence.
ACCESSION BF726110
VERSION BF726110.1 GI:12042021
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 405)
AUTHORS Wistow,G.J., Bernstein,S., Behal,A.G. and Smith,D.
TITLE NEIBANK: EST analysis and bioinformatics for ocular genomics
JOURNAL Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
COMMENT Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: griaeme@helix.nih.gov
Plate: 01 row: c column: 01
Seq primer: -21M13 forward primer (ABI).
Location/Qualifiers
1. 405
/organism="Homo sapiens"

/db_xref="taxon:9606"
/clone="by01c01"
/clone_lib="Human lens cDNA (Un-normalized, unamplified):
By"
/tissue_type="lens"
/dev_stage="Adult"
/lab_host="EMDH10B"
/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
from different adults (both approximately 40 years old)
together yielded 20ug of total RNA and 150ng mRNA for cDNA
library synthesis. A directionally cloned cDNA library in
the pCMVSPORT6 vector was constructed at Life Technologies
, essentially following the protocols of the Superscript
plasmid system full details of which are contained in the
manufacturer's instruction manual
(http://www.lifetech.com/). First strand synthesis was
carried out using a Not I primer-adaptor
[5'-pGACTAGTTCGATGACGACGCGCCGCTT)15-3']. Not I/plunt
end inserts were cloned into the Not I/EcoR V sites in the
vector. EST analysis was performed on the unamplified
library at the NIH Intramural Sequencing Center (NISC)."
BASE COUNT 140 a 58 c 47 g 160 t
ORIGIN

Query Match 60.0%; Score 15; DB 168; Length 405;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 caagtcacagaatta 25
|||||
Db 249 CAAGTCACAGAATTA 263

RESULT 26
BG413230 406 bp mRNA EST 13-MAR-2001
LOCUS ETESTed18903.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
DEFINITION tenella cDNA 5', similar to SW:FBM1_BOVIN P98133 FIBRILIN 1
PRECURSOR ; mRNA sequence.
ACCESSION BG413230
VERSION BG413230.1 GI:13318783
KEYWORDS EST.
SOURCE Eimeria tenella.
ORGANISM Eimeria tenella.
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
Eimeria.
REFERENCE 1 (bases 1 to 406)
AUTHORS Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
Martin,J., Wylie,T., Underwood,K., Stepcoe,M., Theising,B., Allen
,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
,N., Schuck,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
WashU-Merck Eimeria tenella project
Unpublished (1999)
CONTACT: David Sibley, Ph.D.
WashU-Merck Eimeria tenella project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
information relating to organism, libraries, or clone availability.
Seq primer: -400P from Gibco.
Location/Qualifiers
1. 406
/organism="Eimeria tenella"
/strain="LS18"
/db_xref="taxon:5802"
/clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
/dev_stage="Sporozoite stage"
/lab_host="SOLR"

FEATURES
source
1. 405
/organism="Homo sapiens"

/note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
; Sporozoites were obtained from in vitro sporulated and
excysted oocysts of *E. tenella* grown in chickens. cDNA was
synthesized from poly mRNA using an oligo-dT primer
containing a XhoI site. Following second strand synthesis,
EcoRI adaptors were ligated to the cDNA and products were
size-selected on sephacryl S500. The cDNA were ligated to
EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were
converted to phagemids by mass excision using ExAssist
helper phage and *E. coli* SOLR cell (Stratagene). Clones
were selected by negative hybridization against a pool of
overrepresented ESTs (N>10, from 1682 previous reads).
Insert sizes range from 1.2-2.9kb. The library may contain
a small percentage of host or bacterial contaminants."

BASE COUNT

91 a 126 c 114 g 75 t

ORIGIN

Query Match 60.0%; Score 15; DB 153; Length 406;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacatgtcgaact 15
|||||

Db 406 GACATGTCCGAAGT 392

RESULT 27

AI942464 416 bp mRNA EST 06-SEP-1999

LOCUS

wp80c05.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2468072 3',

DEFINITION

mRNA sequence.

ACCESSION

AI942464 GI:5707120

VERSION

EST.

KEYWORDS

human.

SOURCE

human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

1 (bases 1 to 416)

AUTHORS

NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE

National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BRNAP), Tumor Gene Index

JOURNAL

Unpublished (1998)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgaps-rt@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldi, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLT at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco

High quality sequence stop: 410.

FEATURES

source

1. 416

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2468072"

/clone_lib="NCI-CGAP_Brn25"

/tissue_type="anaplastic oligodendroglioma"

/lab_host="DH10B"

/note="Organ: brain; Vector: pTZ19D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTCACCAATCTCAAGTGGAGCGCGCATAGTTTATTTTATTTTATTTTATTTT

T 3'] double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of the modified pTZ19D vector.

Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldi."

BASE COUNT

164 a 63 c 73 g 116 t

ORIGIN

Query Match 60.0%; Score 15; DB 104; Length 416;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaatta 25
|||||

Db 331 CAAGTCACAGAATTA 345

RESULT 28

LOCUS

AA172188 420 bp mRNA EST 09-MAR-1998

DEFINITION

zp29f06.s1 Stratagene neuroepithelium (#937231) Homo sapiens cDNA
clone IMAGE:610883 3', mRNA sequence.

ACCESSION

AA172188 GI:1751265

VERSION

EST.

KEYWORDS

human.

SOURCE

human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

1 (bases 1 to 420)

AUTHORS

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Kizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Matra, M., Martin
J., Moore, B., Schellenberg, K., Stepien, M., Tan, F., Theisling, B.,
White, Y., Wylie, T., Waterston, R. and Wilson, R.

TITLE

WashU-NCI human EST project

JOURNAL

Unpublished (1997)

COMMENT

Contact: Wilison RK

JOURNAL

Washington University School of Medicine

COMMENT

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

COMMENT

Tel: 314 286 1800

COMMENT

Fax: 314 286 1810

COMMENT

Email: estwatson.wustl.edu

COMMENT

This clone is available royalty-free through LNLT; contact the

COMMENT

IMAGE Consortium (info@image.llnl.gov) for further information.

COMMENT

Insert length: 934 Std Error: 0.00

COMMENT

Seq primer: -40M13 fwd. from Amersham

COMMENT

High quality sequence stop: 294.

FEATURES

source

1. 420

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/db_xref="GDB:4626142"

/clone="IMAGE:610883"

/clone_lib="Stratagene neuroepithelium (#937231)"

/dev_stage="Ntera-2/RA neuroepithelial cells"

/lab_host="SOLR (Xenopus resistant)"

/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dT. NT2

cells (Ntera-2/cl.D1) induced with Retinoic Acid for 24
hours. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5'

adaptor sequence: 5' GATTCGCGCAGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTATTTTATTTTATTTT 3' "

sequence: 5' CTCGAGTTTATTTTATTTTATTTT 3' "

sequence: 5' CTCGAGTTTATTTTATTTTATTTT 3' "

sequence: 5' CTCGAGTTTATTTTATTTTATTTT 3' "

sequence: 5' CTCGAGTTTATTTTATTTTATTTT 3' "

sequence: 5' CTCGAGTTTATTTTATTTTATTTT 3' "

sequence: 5' CTCGAGTTTATTTTATTTTATTTT 3' "

sequence: 5' CTCGAGTTTATTTTATTTTATTTT 3' "

sequence: 5' CTCGAGTTTATTTTATTTTATTTT 3' "

sequence: 5' CTCGAGTTTATTTTATTTTATTTT 3' "

sequence: 5' CTCGAGTTTATTTTATTTTATTTT 3' "

sequence: 5' CTCGAGTTTATTTTATTTTATTTT 3' "

sequence: 5' CTCGAGTTTATTTTATTTTATTTT 3' "

sequence: 5' CTCGAGTTTATTTTATTTTATTTT 3' "

sequence: 5' CTCGAGTTTATTTTATTTTATTTT 3' "

sequence: 5' CTCGAGTTTATTTTATTTTATTTT 3' "

sequence: 5' CTCGAGTTTATTTTATTTTATTTT 3' "

sequence: 5' CTCGAGTTTATTTTATTTTATTTT 3' "

sequence: 5' CTCGAGTTTATTTTATTTTATTTT 3' "

sequence: 5' CTCGAGTTTATTTTATTTTATTTT 3' "

sequence: 5' CTCGAGTTTATTTTATTTTATTTT 3' "

```

RESULT 29
A0142435 427 bp DNA GSS 24-SEP-1998
LOCUS HS_3064.A1 All MR Clt Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=3064 Col=21 Row=A, DNA sequence.
ACCESSION A0142435
VERSION A0142435.1 GI:3533088
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 427)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3064 row: A column: 21
Class: BAC ends
High quality sequence stop: 427.
FEATURES
Source Location/Qualifiers
1..427
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Plate=3064 Col=21 Row=A"
/clone_lib="Clit Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelBAC11; BAC Clones in
E-coli DH10B"
BASE COUNT 137 a 78 c 103 g 108 t 1 others
ORIGIN
Query Match 60.0%; Score 15; DB 224; Length 427;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 gcaagtcacagaatt 24
|||||
Db 198 GCAAGTCACAGAAAT 184

RESULT 30
A0870882 429 bp mRNA EST 05-APR-2001
LOCUS ras1b09.y1 Bird-Rao Meloidogyne incognita J2 Meloidogyne incognita
DEFINITION CDNA 5' similar to WP:C26D10.4 CE03027 ;, mRNA sequence.
ACCESSION A0870882
VERSION A0870882.1 GI:8004935
KEYWORDS EST.
SOURCE southern root-knot nematode.
ORGANISM Meloidogyne incognita
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchima;
Tylenchoidae; Heteroderidae; Meloidogyinae; Meloidogyne.
REFERENCE 1 (bases 1 to 429)
AUTHORS McCarter,J., Clifton,S., Chippelli,B., Pape,D., Martin,J., Wylie,T.,
Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
Gibbons,M., Rittler,E., Bennett,J., Franklin,C., Tsagarashvili,R.,
Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
,M., Allen,M., Person,B., Swalter,T., Harvey,N., Schurk,R., Kohn,S.,
Shin,T., Jackson,Y., Cardenas,M., McCanm,R., Waterston,R. and
Wilson,R.

```

```

TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
The library was constructed by Uma Rao and David Bird
(david_bird@wustl.wustl.edu) at North Carolina State University. DNA
Sequencing by: Washington University Genome Sequencing Center St.
Louis.
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 428.
FEATURES
Source Location/Qualifiers
1..429
/organism="Meloidogyne incognita"
/db_xref="taxon:6306"
/clone_lib="Bird-Rao Meloidogyne incognita J2"
/dev_stage="enriched for 2nd stage juveniles"
/lab_host="XJ0RL"
/note="Vector: ZAP express - PBKCMV (Stratagene); Site_1:
EcoRI; Site_2: XhoI; Oligo (dT) primed library. cDNA was
constructed and cloned unidirectionally into the vector
within the 5' EcoRI and 3' XhoI sites. This library was
constructed by Dr. Uma Rao and Dr. David Bird at North
Carolina state University"
BASE COUNT 127 a 66 c 38 g 198 t
ORIGIN
Query Match 60.0%; Score 15; DB 121; Length 429;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 caagtcacagaatta 25
|||||
Db 385 CAAAGTCACAGAAATTA 371

RESULT 31
A0086771 454 bp DNA GSS 26-AUG-1998
LOCUS HS_2175.B2.A06-MR Clt Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=2175 Col=12 Row=B, DNA sequence.
ACCESSION A0086771
VERSION A0086771.1 GI:3455988
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 454)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2175 row: B column: 12
Class: BAC ends
High quality sequence stop: 454.

```

FEATURES

Location/Qualifiers
1. .454

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Plate-2175 Col-12 Row-B"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 188 a 70 c 76 g 117 t 3 others
ORIGIN

Query Match 60.0%; Score 15; DB 224; Length 454;
Best Local Similarity 100.0%; Pred. NO. 85;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caagtccagaatta 25
11 caagtccagaatta 170

RESULT 32
AA973493 455 bp mRNA EST 17-JUN-1998
LOCUS
DEFINITION O045605.S1 NCI_CGAP_Lu5 Homo sapiens CDNA clone IMAGE:1569152 3',
mRNA sequence.
ACCESSION AA973493
VERSION AA973493.1 GI:3148673
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 455)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www-bio.llnl.gov/bdrrp/image/image.html
Insert length: 553 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 440.
Location/Qualifiers

FEATURES
source 1. .455

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1569152"
/clone_lib="NCI_CGAP_Lu5"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker: 1st strand CDNA was prepared from a
neuroendocrine lung carcinoid, and was then primed with a
Not I - oligo(dT) primer. Double-stranded CDNA was ligated
to Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library is normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 175 a 64 c 84 g 132 t
ORIGIN

Query Match 60.0%; Score 15; DB 14; Length 455;

Best Local Similarity 100.0%; Pred. NO. 85;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caagtccagaatta 25
11 caagtccagaatta 354

RESULT 33
A1756736 456 bp mRNA EST 18-JAN-2000
LOCUS
DEFINITION ESTE2a24h07.y1 Eimeria S5-2 Sporozoite stage Eimeria tenella CDNA
5' similar to SW:FBW1_BOVIN P98133 FIBRILLIN 1 PRECURSOR ; mRNA
sequence.
ACCESSION A1756736
VERSION A1756736.1 GI:5150459
KEYWORDS EST.
SOURCE Eimeria tenella.
ORGANISM Eimeria tenella.
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
Eimeria.

REFERENCE 1 (bases 1 to 456)
Liberator, P., Diaz, C., Tang, K., Marra, M., Hillier, L., Kucaba, T.,
Martin, J., Wyllie, T., Underwood, K., Steptoe, M., Theising, B., Allen
M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey
N., Schurk, R., Ritter, E., Kohn, S., Florence, N., Shin, T., Jackson
Y., Cardenas, M., McCann, R., Waterson, R., Wilson, R. and Sibley, D.
Washu-Merck Eimeria tenella project
Unpublished (1999)
CONTACT: David Sibley, Ph.D.
Washu-Merck Eimeria tenella project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Contact David Sibley (toxoestporcim.wustl.edu) for further
information relating to organism, libraries, or clone availability.
Seq primer: -40RP from Gibco
High quality sequence stop: 419.
Location/Qualifiers

FEATURES
source 1. .456

/organism="Eimeria tenella"
/strain="LS18"
/db_xref="taxon:5802"
/clone_lib="Eimeria S5-2 Sporozoite stage"
/dev_stage="Sporozoite"
/lab_host="SOLR E. coli"
/note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
; Sporozoites were obtained from in vitro sporulated and
excysted oocysts of E. tenella grown in chickens. CDNA
was synthesized from poly mRNA using an oligo-dT primer
containing a XhoI site. Following second strand synthesis,
EcoRI adaptors were ligated to the cDNA and products were
size-selected on Sephacryl S500. cDNAs were digested with
EcoRI/XhoI and cloned into lambda Zap II (Stratagene).
Clones were converted to phagemids by mass excision using
EXassist helper phage and SOLR cells (Stratagene).
Insert sizes range from 1.2-2.9 kb."

BASE COUNT 100 a 137 c 132 g 87 t
ORIGIN

Query Match 60.0%; Score 15; DB 24; Length 456;
Best Local Similarity 100.0%; Pred. NO. 85;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gacattgcgaact 15
11 gacattgcgaact 421

RESULT 34

Email: cgapbs-remail.nih.gov
Life Technologies catalog #: 11548-013

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bdrp/image/image.html

Insert length: 1533 Std Error: 0.00

Seq primer: -400P from Gibco

High quality sequence stop: 399.
Location/Qualifiers

1. 471

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2294845"

/clone_lib="NCI-CGAP_Pant"

/tissue_type="adenocarcinoma"

/note="Organ: pancreas; Vector: PCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.72 kb. Life Technologies catalog #:
11548-013"

BASE COUNT 171 a 69 c 53 g 178 t

ORIGIN

Query Match 60.0%; Score 15; DB 22; Length 471;

Best Local Similarity 100.0%; Pred. No. 85;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 caagtcacagaatta 25

Db 235 CAAGTCACAGATT 249

RESULT 37

LOCUS BG020076

DEFINITION

3 similar to TR:Q9VCA6 Q9VCA6 CG6313 PROTEIN.; mRNA sequence.

ACCESSION BG020076

VERSION BG020076.1

KEYWORDS

SOURCE

ORGANISM

EST.

Xenopus laevis

African clawed frog.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

Xenopodinae; Xenopus.

1 (bases 1 to 485)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: Xenopus clones from this library are available

through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov

Seq primer: -400P from Gibco

High quality sequence stop: 409.

Location/Qualifiers

1. 485

/organism="Xenopus laevis"

/db_xref="taxon:8355"

/clone="IMAGE:3400308"

/clone_lib="NCI-CGAP_Pant"

/tissue_type="embryo (stages 24-25)"

/lab_host="DH10B (phage-resistant)"

/note="Vector: PCMV-SPORT6; Site_1: NotI; Site_2: SalI;
Cloned unidirectionally. Primer: Oligo dT. Average insert
size 1.7 kb. Constructed by Life Technologies. Note: This

BASE COUNT 117 a 102 c 91 g 165 t 10 others

ORIGIN

Query Match 60.0%; Score 15; DB 172; Length 485;

Best Local Similarity 100.0%; Pred. No. 85;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gcaagtcacagaatt 24

Db 437 GCAAGTCACAGATT 451

RESULT 38

LOCUS BG516037

DEFINITION

ESTED48c02.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria

tenella cDNA 5' similar to SW:FBNI_BOVIN P98133 FIBRILLIN 1

PRECURSOR.; mRNA sequence.

ACCESSION BG516037

VERSION BG516037.1

KEYWORDS

SOURCE

ORGANISM

EST.

Eimeria tenella.

Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;

Eimeria.

1 (bases 1 to 487)

Liberator, P., Diaz, C., Tang, K., Maria, M., Hillier, L., Kucaba, T.,

Martin, J., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen

M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey

N., Schurk, R., Ritter, E., Kohn, S., Florence, N., Shin, T., Jackson

, Y., Cardenas, M., McCann, R., Waterston, R., Wilson, R. and Sibley, D.

Washu-Merck Eimeria tenella project

Unpublished (1999)

Contact: David Sibley, Ph.D.

Washu-Merck Eimeria tenella project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: estelw@wustl.edu

Contact David Sibley (toxest@porcim.wustl.edu) for further

information relating to organism, libraries, or clone availability.

Seq primer: -40RP from Gibco

High quality sequence stop: 403.

Location/Qualifiers

1. 487

/organism="Eimeria tenella"

/strain="LS18"

/db_xref="taxon:5802"

/clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"

/dev_stage="Sporozoite stage"

/lab_host="SOLR"

/note="Vector: Bluescript SK-; Site_1: EcoRI, site_2: XhoI

: Sporozoites were obtained from in vitro sporulated and

excysted oocysts of E. tenella grown in chickens. cDNA was

synthesized from poly mRNA using an oligo-dT primer

containing a XhoI site. Following second strand synthesis,

EcoRI adaptors were ligated to the cDNA and products were

size-selected on sephacryl 500. The cDNA were ligated to

EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were

converted to phagmids by mass excision using EXAssist

helper phage and E.coli SOLR cell (Stratagene). Clones

were selected by negative hybridization against a pool of

overrepresented ESTs (N=10, from 1982 previous reads).

Insert sizes range from 1.2-2.9kb. The library may contain

a small percentage of host or bacterial contaminants."

BASE COUNT 105 a 147 c 142 g 93 t

ORIGIN

Query Match 60.0%; Score 15; DB 154; Length 487;

Best Local Similarity 100.0%; Pred. No. 85;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gacatttcgaagt 15
|||||

Db 451 GACATTTCGCAAGT 437

RESULT 39
BF775935 489 bp mRNA EST 12-JAN-2001
LOCUS
DEFINITION 286217 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BF775935
VERSION BF775935.1 GI:12123835
KEYWORDS EST.
SOURCE
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE
AUTHORS 1 (bases 1 to 489)
Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
W.M. and Keele,J.W.
Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
Unpublished (2000)
TITILE
JOURNAL
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail.marc.usda.gov
Single base sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross-match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAAT
BACKWARD: GTTTCACAGTCAGCAGC
Plate: 88 row: K column: 20
Seq primer: ATTGAGCGACATATGAC.
FEATURES
Source
Location/Qualifiers
1..489
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."
BASE COUNT 147 a 85 c 95 g 162 t
ORIGIN

Query Match 60.0%; Score 15; DB 169; Length 489;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 11 caagtcacagaatt 25
|||||

Db 153 CAAGTCACAGAATT 139

RESULT 40
A2398330 498 bp DNA GSS 03-OCT-2000
LOCUS
DEFINITION IM0165N1R Mouse 10Kb plasmid UUGC1M library Mus musculus genomic
ACCESSION A2398330
VERSION A2398330.1 GI:10513498
KEYWORDS GSS.

SOURCE
ORGANISM house mouse.
Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 498)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10Kb
plasmid inserts

TITLE
JOURNAL
COMMENT Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0163 row: N column: 15
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 498.

FEATURES
Source
Location/Qualifiers
1..498
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="UUGC1M0163N15"
/clone_lib="Mouse 10Kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/note="Vector: pMD22nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (q114732114|9b|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 187 a 91 c 117 g 103 t
ORIGIN

Query Match 60.0%; Score 15; DB 243; Length 498;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 gcaagtcacagaatt 24
|||||

Db 101 GCAAGTCACAGAATT 115

RESULT 41
AA846453 502 bp mRNA EST 31-DEC-1998
LOCUS
DEFINITION aj56b03.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1394285
ACCESSION AA846453
VERSION AA846453.1 GI:2932593

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 502)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrp/image/image.html
Insert Length: 925 Std Error: 0.00
Seq primer: -40mj3 fwd. ET from Amersham
High quality sequence stop: 448.
FEATURES
Location/Qualifiers
1..502
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1394285"
/clone_lib="Soares_testis_NHT"
/lab_host="DH10B"
/sex="male"
/note="Vector: p7T73D-Pac (Pharmacia) with a modified
polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I-oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGCCGCCAATTTTCTTTTCTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T73 vector. Library
went through one round of normalization to cots5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 196 a 67 c 97 g 142 t
ORIGIN

Query Match 60.0%; Score 15; DB 12; Length 502;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaacta 25
|||||

Db 342 CAAGTCACAGAAATTA 356

RESULT 42
AI672463 502 bp mRNA EST 16-DEC-1999
LOCUS AI672463
DEFINITION wa03e05.x1 NCI-CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2297024 3',
mRNA sequence.
ACCESSION AI672463
VERSION AI672463.1 GI:4852194
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 502)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrp/image/image.html
Insert Length: 535 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 458.
FEATURES
Location/Qualifiers
1..502
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2297024"
/clone_lib="NCI CGAP_Kid11"
/lab_host="DH10B"
/note="Organ: kidney; Vector: p7T73D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI;
plasmid DNA from the normalized library NCI-CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneids 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."

BASE COUNT 198 a 67 c 96 g 141 t
ORIGIN

Query Match 60.0%; Score 15; DB 23; Length 502;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 caagtcacagaacta 25
|||||

Db 339 CAAGTCACAGAAATTA 353

RESULT 43
AI634722 503 bp mRNA EST 16-DEC-1999
LOCUS AI634722
DEFINITION tz09e04.x1 NCI-CGAP_U11 Homo sapiens cDNA clone IMAGE:2288094 3',
similar to contains element XTR repetitive element ;, mRNA
sequence.
ACCESSION AI634722
VERSION AI634722.1 GI:4686052
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 503)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrp/image/image.html
Insert Length: 3719 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 413.
FEATURES
Location/Qualifiers
1..503

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:228094"
/clone_lib="NCI-CGAP_Ut1"
/tissue_type="well-differentiated endometrial
adenocarcinoma, 7 pooled tumors"
/lab_host="DH10B"
/Note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Life Technologies catalog #:
11538-014"

BASE COUNT      184 a      77 c      56 g      186 t
ORIGIN

Query Match      60.0%; Score 15; DB 22; Length 503;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      11 caagtcacagaatt 25
Db      236 CAAGTCACAGAATT 250

RESULT  44
AM418528      512 bp      mRNA      EST      09-FEB-2000
LOCUS      xr32c07.x1 NCI-CGAP_Ut4 Homo sapiens cDNA clone IMAGE:2761836 3'
DEFINITION      similar to contains element MER32 repetitive element ;, mRNA
SEQUENCE
AM418528.1 GI:6946410
EST.
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 512)
REFERENCE      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE      Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:
www.bio.linnl.gov/bhrp/image/image.html
Seq primer: -400p from Gibco
High quality sequence stop: 404.
Location/Qualifiers
1..512
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2761836"
/clone_lib="NCI-CGAP_Ut4"
/tissue_type="serous papillary carcinoma, high grade, 2
pooled tumors"
/lab_host="DH10B"
/Note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.48 kb. Life Technologies catalog #:
11542-016"

BASE COUNT      186 a      79 c      56 g      191 t
ORIGIN

Query Match      60.0%; Score 15; DB 115; Length 512;
Best Local Similarity 100.0%; Pred. No. 86;

```

```

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      11 caagtcacagaatt 25
Db      240 CAAGTCACAGAATT 254

RESULT  45
AO801291      512 bp      DNA      GSS      09-AUG-1999
LOCUS      HS-2088.A1.D05.T7C CIT Approved Human Genomic Sperm Library D Homo
DEFINITION      sapiens genomic clone Plate-2088 Col=9 Row=G, DNA sequence.
ACCESSION      AO801291 GI:5718623
VERSION
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 512)
REFERENCE      Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
AUTHORS      Mahairas,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE      Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL      Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE      99380589
COMMENT      Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web server: http://www.htsc.washington.edu
Plate: 2088 row: G column: 9
Seq primer: T7
Class: BAC ends
High quality sequence stop: 512.
Location/Qualifiers
1..512
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-2088 Col=9 Row=G"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/Note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
E-coli DH10B"

BASE COUNT      144 a      117 c      96 g      144 t      11 others
ORIGIN

Query Match      60.0%; Score 15; DB 233; Length 512;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

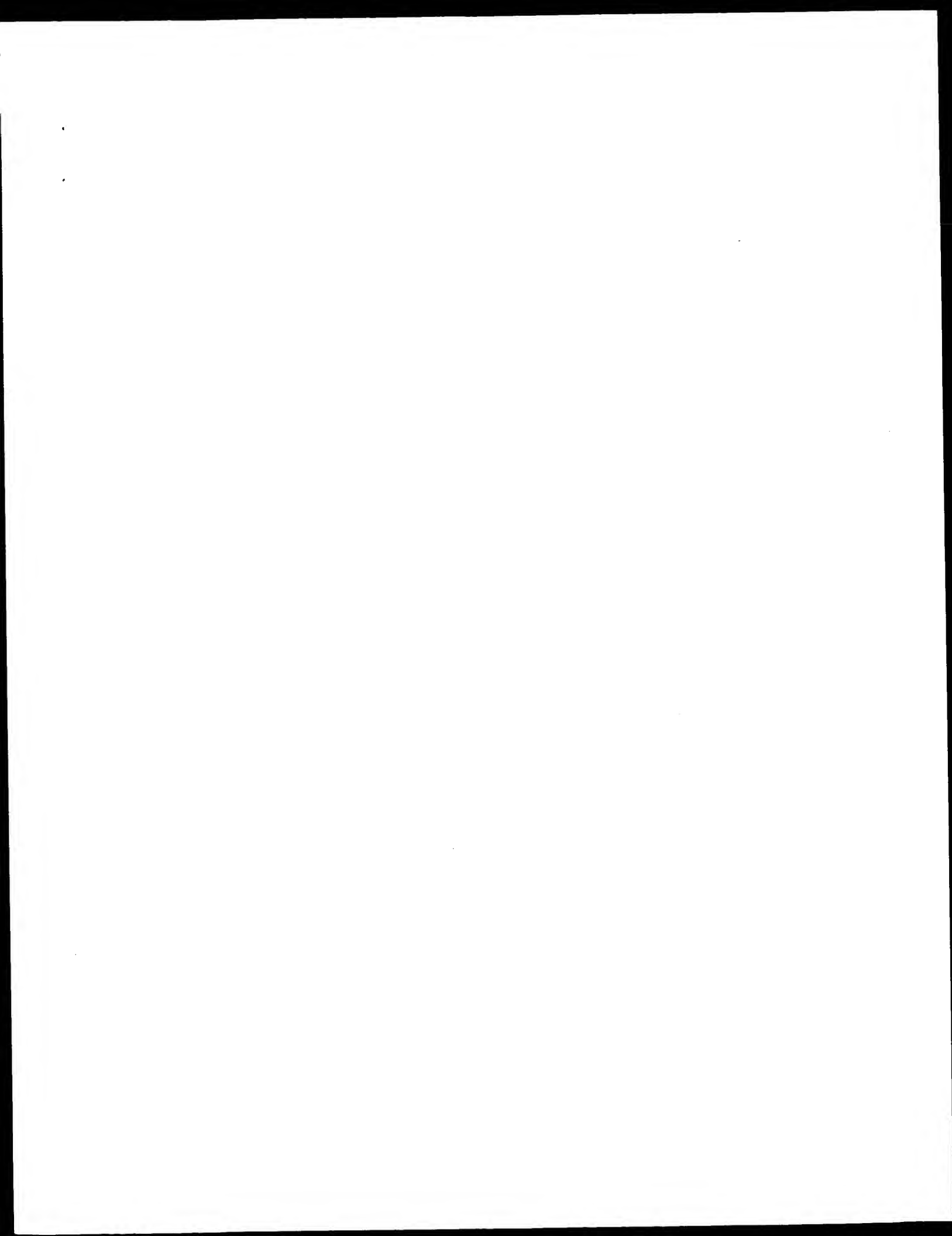
QY      10 gcaagtcacagaatt 24
Db      248 GCAAGTCACAGAATT 262

Search completed: October 9, 2001, 15:15:21
Job time: 13643 sec

```

Wed Oct 10 07:45:33 2001

us-09-396-196f-1.01i.rst



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:48:27 ; Search time 1670.83 Seconds
(without alignments)
231.438 Million cell updates/sec

Title: US-09-396-196f-10
Perfect score: 25
Sequence: 1 tgcacggtgcagtcacgacgttg 25

Scoring table: OLIGO_NTC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 773874588 residues

Word size : 9

Total number of hits satisfying chosen parameters: 210607

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_om:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rnd:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rnd:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vl:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vl1:*
59: gb_vl2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	9 AR029499	AR029499 Sequence
2	25	100.0	1041	9 AR034916	AR034916 Sequence
3	25	100.0	1084	9 A11530	A11530 BioB gene o
4	25	100.0	1121	10 E00893	E00893 Genomic DNA
5	25	100.0	5793	2 E00893	J04423 E.coli 7,8-
6	25	100.0	5872	9 A38246	A38246 Sequence 1
7	25	100.0	5872	9 A38251	A38251 Sequence 6
8	25	100.0	5872	9 A93674	A93674 Sequence 1

```

9      25 100.0 5872 9 A93679 Sequence 6
10     25 100.0 5872 9 AR101809 Sequence
11     25 100.0 5872 9 AR101810 Sequence
12     25 100.0 11022 9 AR000180 Sequence
13     21 84.0 7215 3 SMABIO
14     21 84.0 13501 1 AR005258
15     21 84.0 297816 2 AR002553
16     17 76.0 213732 1 AE001862
17     17 76.0 2081 94 HAMHP
18     17 68.0 7318 94 AB026497
19     17 68.0 69900 66 AC021767
20     17 68.0 151340 74 AC069132
21     17 68.0 155204 73 AC068100
22     17 68.0 168360 71 AC041048
23     17 68.0 176379 72 AC060833
24     17 68.0 200729 67 AC022781
25     16 64.0 1044 53 CNS07873
26     16 64.0 8326 8 AE237379
27     16 64.0 10536 1 AE004050
28     16 64.0 12242 1 AE005241
29     16 64.0 38065 3 MCB1450
30     16 64.0 39520 78 AC091120
31     16 64.0 71230 85 AC004508
32     16 64.0 85633 85 AC002088
33     16 64.0 109973 60 AC007555
34     16 64.0 110000 84 HSMX1.4
35     16 64.0 147129 74 AC073052
36     16 64.0 155628 71 AC027810
37     16 64.0 156119 74 AC073247
38     16 64.0 162043 71 AC040930
39     16 64.0 175459 73 AC068538
40     16 64.0 183700 86 AC005412
41     16 64.0 185492 69 AC025339
42     16 64.0 202828 72 AC064835
43     16 64.0 245489 64 AC016186
44     16 64.0 262278 93 AP002552
45     16 64.0 313064 93 HSMX1B

```

ALIGNMENTS

```

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton, D. Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
    source location/Qualifiers
        1..1041
        /organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 tcgtcaggtcaggtcaggtcaggtg 25
|||||
Db 111 TCGTCAGGTGACGTCAGCAGCTTG 135

```

RESULT 2

```

AR034916
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton, D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
    source location/Qualifiers
        1..1041
        /organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 tcgtcaggtcaggtcaggtcaggtg 25
|||||
Db 111 TCGTCAGGTGACGTCAGCAGCTTG 135

```

```

RESULT 3
LOCUS A11530 1084 bp DNA
DEFINITION B10B gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 1084)
JOURNAL Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES
    source location/Qualifiers
        1..1084
        /organism="Escherichia coli"
        /db_xref="taxon:562"
        24..1064
        /gene="B10B"
        24..1064
        /gene="B10B"
        /codon_start=1
        /transl_table=11
        /protein_id="CAA00965.1"
        /db_xref="GI:490219"
        /db_xref="SWISS-PROT:P12996"
        /translation="MAHRPMTLSOYTELFEKPLDILFEAQOVHQRHPPQVOYST
        LLSIKTGACPEDECKYCPQSSRYTGTGLAEKRLMEVEYLESARKARAGSTRQMGAW
        KNPHERDMPLEBVOGVKRAMGLEACMTIGTISESOAKRLANAGLDYVNHNLDTSPF
        VPINLVKVGTEPLADNDVDVADFDITRIYAVRIMMPTSVYRLSAGREOMNQTQAMC
        FMAGANSIFGCKLITTPNEEDKDLQLEFRKILGNQAVAVLAGDNEQQRLEQALMT
        PPDDEVYNAAL"
        CDS

```

```

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 9; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 tcgtcaggtcaggtcaggtcaggtg 25
|||||

```


Db 134 TCGTACGTCAGTCAGCAGCTTG 158

RESULT 4

LOCUS E00893 1121 bp DNA PAT 29-SEP-1997

DEFINITION Genomic DNA encoding biotin Synthetase.

ACCSSION E00893

VERSION E00893.1 GI:2169154

KEYWORDS JP 1986149091-A/1.

SOURCE Escherichia coli.

ORGANISM Escherichia coli.

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 1121)

AUTHORS Hirono, Y., Kojima, T. and Kimura, H.

TITLE HIRONO, YOSHIIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC

JOURNAL PRODUCTION OF BIOTIN

DIPLIX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND

PATENT: JP 1986149091-A 1 07-JUL-1986;

COMMENT NIPPON SODA CO LTD

OS Escherichia coli

PN JP 1986149091-A/1

PD 07-JUL-1986

PF 24-DEC-1984 JP 1984272605

PI HIRONO, YOSHIIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC

C12N15/00, C12N1/20, C12P13/18, (C12N1/20, C12R1:19), (C12P13/18, PC

C12R1:19);

CC strandedness: Double;

CC topology: Linear;

CC hypothetical: No;

CC anti-sense: No;

CC *source: strain-Escherichia coli Ns101;

CC Feature is identified by experimental;

FH Key Location/Qualifiers

FH FT CDS 42..1079

FEATURES

source Location/Qualifiers

1..1121

/organism="Escherichia coli"

/db_xref="taxon:562"

BASE COUNT 289 a 296 c 325 g 211 t

ORIGIN

Query Match 100.0%; Score 25; DB 10; Length 1121;

Best Local Similarity 100.0%; Pred. No. 0.0027;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgtcaggtcagtcagtcagtcagtcg 25

Db 152 TCGTACGTCAGTCAGTCAGCAGCTTG 176

RESULT 5

LOCUS E00893 5793 bp DNA BCF 28-FEB-1994

DEFINITION E.coli 7,8-diamino-pelargonic acid (bioc), biotin synthetase (biob), 7-keto-8-amino-pelargonic acid synthetase (biof), bioc protein, and dehydrobiotin synthetase (biobd), complete cds.

ACCSSION J04423

VERSION J04423.1 GI:145422

KEYWORDS 7,8-diamino-pelargonic acid aminotransferase; 7-keto-8-amino-pelargonic acid synthetase; bioc gene; biob gene; bioc gene; biob gene; biof gene; biotin synthetase; dehydrobiotin synthetase.

SOURCE Escherichia coli (strain K-12) DNA.

ORGANISM Escherichia coli

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 5793)

AUTHORS Otsuka, A.J., Buomcrislian, M.R., Howard, P.K., Flamm, J. and Johnson, O.

TITLE The Escherichia coli biotin biosynthetic enzyme sequences

JOURNAL J. Biol. Chem. 263, 19577-19585 (1988)

COMMENT 89066784

COMMENT Draft entry and computer-readable sequence [1] kindly submitted by A.Otsuka, 09-NOV-1988.

FEATURES

source

gene

CDS

1..5793

/organism="Escherichia coli"

/strain="K-12"

/db_xref="taxon:562"

complement(98..574)

/note="ORF 1"

/codon_start=1

/transl_table=11

/protein_id="AAA23513.1"

/db_xref="GI:453168"

/translation="MKLISNDLRDGDKLPHRHVNGMGVDGDNISPHLADDPAGTK SFVVCYDPPAPYSGMHHVYVNLPPADRVLPQGGSLVAMPDGVLOTRTDGGKTG YDGAAPPKGTHRYFTVHALDIERIDVDEGASAAAVGFVHFSLASASITAFS"

complement(633..1925)

/gene="biob"

complement(633..1925)

/codon_start=1

/gene="biob"

/product="7,8-diamino-pelargonic acid aminotransferase"

/transl_table=11

/protein_id="AAA23514.1"

/db_xref="GI:457106"

/translation="MTTDLAPDQRIHWPYSMTSPLPYVPVVSNEGCEILISNGR ALVDGSMWMAAIGHYVNHNPOLNAAMKSDIDAMSHVMFGITPAFIELCRKLYAMGRN ALBCEVFLADGSVAVEVAMKMALOYQAGKARORPLTFRNQYHSDTFKAMSCVDPN SMHSIMKGYLPENLFPAPQSRMDGEMDERDYGAFARLMAARHETIAVTEPIYGA GGMVYHPENLKRIRKICOREGILLIADETAGFGRTGLFCEHAEIADILLICGKA LTGGTMTLSATTTREYAEITISGECAGCEMGPTEGPNLACAAASLALIESGDMQ QOVADIEVOLREODLPARDAEMVADVIVGALIGVETTPYVMAALOKRFFVGWIR PFGKLIYMPYIILLPOOLRLTAAVNRAVODETFECQ"

2012..3052

/gene="biob"

2012..3052

/gene="biob"

/codon_start=1

/transl_table=11

/product="biotin synthetase"

/protein_id="AAA23515.1"

/db_xref="GI:145425"

/translation="MAHRPRWTLISQVTELPFKPLDLLPEAOOVHROHFPDPOVST LLSRTGACPEPDCYKCYOTSRYKTGLEAEKLEMEVOYTESARAKAAGSRRCMGAW KNPHRDMPLYLEOVGVKAMGLEACMTTLTSLSESQORLANNGIDYVNHNDTSPF YGNITITRTVOERLDTLEKVRDAGIKVCSGIVGLGTYVDRAGLLDLANLPTPES VPINMLVVKGTPLADNDVDVADFPRTTIVARIIMPTSVVRLSAGREOMNEOTAMC FMAGANSIFVCGKLLTTPNEEDKDLQLFKRLGLNPOQTVAVLAGDNEODQRLQALMT EPDEYVMAAL"

3049..4203

/gene="biob"

3049..4203

/gene="biob"

/codon_start=1

/transl_table=11

/product="7-keto-8-amino pelargonic acid synthetase"

/protein_id="AAA23516.1"

/db_xref="GI:145426"

/translation="MSWQEKINAAIDARRAADALRRYPVACGAGRMVADVDDQYLNF SNDYLGLSHHPDILRAMQOAGDEFGISGSGSVASVYVHQALEELAELEWGYSSA ILFISGFANAOVIAAMAKEDRIAARLSHSLLEASLSPSOLRRFANNDVTHLAR ILASPCPOOWVTEGVSGMDSAPLAETIOVTOOHNGWMLVDADGTVIEGOGS SCWIDKVPDELLVTFPGKGVSGAAVCSSTVDYLLQFARHLITSTSPAPQAOAL RASLAVINSDEGDARREKLALITFRAGVODLPETLADSCSATIOPLIVDPSRAQL AKRLQOQGWVAIRPVPVAGTARLRLITAAHMODIDRLLEVLIENG"

4190..4945

/gene="bioc"

4190..4945

/gene="bioc"

/gene="bioc"

```

/codon_start=1
/transl_table=1
/protein_id="AA23517.1"
/db_xref="GI:145427"
/translation="MATVYKQATIAAGFRAAHYEQHADLQROSDALLMLPORKKT
HVLDCGCGPGMSRHRMRHQAQYALDLSPPMLVOAKQKADADHYTAGDIESPLATA
TFDLAMSLAYOKNGCNLSTALRELRYVRKGVAFPTLVQSLPERHQAQVADERP
HAKRFLPDELEOSLNGVHYQHIOPTTLWFDLALSMRSLKIGATHLEGRDPRIL
TRSLQRLQALAMPQOQGRPYLTHLFLGVARE"
4938..5597
/gene="biob"
/feature="biob"
4938..5597
/gene="biob"
/codon_start=1
/transl_table=1
/product="dethiobiotin synthetase"
/protein_id="AA23518.1"
/db_xref="GI:145428"
/translation="MSKRYFVGTGDFPVKGTASCALLQAKRAGRTAGTKFVASGS
KPEPEGRNSDALALORNSLIDYATVNTYTTAEPTSPHITSAQEGRPLESLVMSAG
LRALEQADAVLYVEGAGCGFTPLSDFTTFADWVQEQPLVLYVGVKIGCINHAMLTA
QVIOHAGLTLAGVANDVTPPGKRHAEYMTTLRMIPRCWERSPGLOKIOKMRPES
T"

BASE COUNT      1363 a   1554 c   1631 g   1245 t
ORIGIN           4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match      100.0%; Score 25; DB 2; Length 5793;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcacagtcagcagtcagcagtcg 25
|||||
Db 2122 tctcaggtgcagtcgacagcttg 2146

RESULT 6
LOCUS           A38246           5872 bp      DNA
DEFINITION      Sequence 1 from Patent WO9408023.
ACCESSION       A38246
VERSION         A38246.1 GI:2294844
KEYWORDS
SOURCE          Escherichia coli.
ORGANISM        Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
                Escherichia.
REFERENCE       1 (bases 1 to 5872)
AUTHORS        Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE          BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL         Patent: WO 9408023-A I 14-APR-1994;
                LONZA AG (CH)
COMMENT         Other publication PL 308301 950724
                Other publication CA 2145400 940414
                Other publication AU 4820293 940426
                Other publication HU 71781 960228
                Other publication SK 42095 951108
                Other publication CZ 9500809 950913
                Other publication FI 951547 950331
                Other publication JP 8501694T 960227.
                Location/Qualifiers
                1..5872
                /organism="Escherichia coli"
                /strain="DSM498"
                /db_xref="taxon:562"
                /clone="PBO30A-15/9"
                1..96
                /function="PROMOTER PTAC"
                /evidence=experimental
                23..28
                /evidence=experimental
                45..49
                /standard_name="PROMOTER PTAC"
                /evidence=experimental

gene
CDS
105..119
/standard_name="BIOB RBS NO.9"
/evidence=experimental
117..1157
/gene="BIOB"
117..1157
/gene="BIOB"
/codon_start=1
/transl_table=11
/number=1
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAA02325.1"
/db_xref="GI:2294845"
/translation="MAHRPRRTLSQVTELEFEPRLDILFEAQYHROHEDPRQVQST
LTSITGACPEPCKYCPQSSRYKTGLEAEERLMEVQVLESARAKKAAGSTFCGAAW
KNPHEDEMFLEQMVQVYKAMGLACMTLGLTSESQAOQLANAGIDYNNHLDPSPEF
YGNITTTTRYQERLDLTLEKVRDAGIKVSGGIVIGETVKDRAGILLQALMLPPRES
VFINMLVYKGTPLADNDVDVADFIRTIIVARLIMPTSYVRSLSGRQNMVEQIOAMC
FMAGANSIFVGGKILTTTPNEDEKDLQLEFKRLGINTPOOTAVIAGDNQQRLEQALMT
PTEDEYVNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/number=3
/product="PROTEIN"
/protein_id="CAA02326.1"
/db_xref="GI:2294846"
/translation="MATVYKQATIAAGFRAAHYEQHADLQROSDALLMLPORKKT
HVLDCGCGPGMSRHRMRHQAQYALDLSPPMLVOAKQKADADHYTAGDIESPLATA
TFDLAMSLAYOKNGCNLSTALRELRYVRKGVAFPTLVQSLPERHQAQVADERP
HAKRFLPDELEOSLNGVHYQHIOPTTLWFDLALSMRSLKIGATHLEGRDPRIL
TRSLQRLQALAMPQOQGRPYLTHLFLGVARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:
8-AMINO-7-OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/number=5
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAA02327.1"
/db_xref="GI:2294847"
/translation="MTTDLALPDORHIMHPTYSMTSPLYPVYVASEGCELLISDGR
LVCYSSWMAAIGHYVNPOLNAAMKQSIDAMSHVMGTTTHAPALELCRLVAMTPOP
LEVTFLADSSAVAEVAMKMAILOYQAKGARQRLTFENNGHGTGFAAGSYCDPDS
MBSLWNGYLPENLEFAPAPQSRMDGENDERMDWGFARLMAHREHIAVITEPIVQAG
GMRNYHEWMLKRIKICDRGILLIADDEIATGFGRTGLFACEHAIADIIDIGKAL
TGCTMTLSATLITTRVEAETISNGEACFPWGCPTMGNPLACAAANSLAILESGMQO
OVADIEVOLREOLAPARDAEMVADVIGALICVETTHYVNAALOKFVEEGCWIRP
EGCLITLMPYIILPQOOLRLDVAVRVADDETFQO"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/number=6

```

/evidence=experimental
/product="PROTEIN"
/protein_id="CAA02328.1"
/db_xref="GI:2294848"
/translation="MKLISNDLRGDKLPHRHVENGMGVDDGDNISPHLAMDVPAKTK
SPVVCYDPADAPTSQGMWVYVNPADTRVLPQGFSGILVAMPDGVLOTDRGKTKG
YDGAAPPKGCTHRYIFTYHALDIEKIDVDEGASGMVGFNFHFLASASITAMS"

stem_loop
5583. .5605
5583. .5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"

BASE COUNT 1318 a 1552 c 1695 g 1307 t

ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tcgtcaggtcaggtcagcagcttg 25
|||||
Db 227 TCCTCAGGTGTCAGGTGTCAGCAGCTTG 251

RESULT 7
A38251 5872 bp DNA PAT 05-MAR-1997
LOCUS A38251
DEFINITION Sequence 6 from Patent WO9408023.
ACCESSION A38251
VERSION A38251.1 GI:2294849
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL Patent: WO 9408023-A 6 14-APR-1994;
LONZA AG (CH)
COMMENT
Other publication PL 308301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42095 951108
Other publication CZ 9500809 950913
Other publication FI 951547 950331
Other publication JP 8501694T 960227.
location/Qualifiers
1. .5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141. .1156
/standard_name="BIOF RBS"
1154. .2308
/gene="BIOF"
1154. .2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/number=2
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAA02329.1"
/db_xref="GI:2294850"
/translation="MSWQEKINALLARRAARADLRRTYPAQAGRWIVADDDROYLNF
SSNDYGLSHHPDITIRAWQCAQDFGIGSGGSGVSGSYVHQALELELEKWLGYSKA
LFLTSGFAANQAVIAAMAKEDRIIAADRLSHASLEASLSLSPQLRFAINDVTHLAR
LLASPCGQOMVYEGVFSMDGDSAPLAEIQVYTOYHNGWLMWDAGHTGTVIGDGRG

SCMLQKVPPELLVYTFGKGFVGSGAAVLCSTVADYLLQFARHLVYSTMPAQOAL
RASLAVIRSDGDAKREKLAALITREFAGVODLPTLADSCSAIOPLIVGDNRAJOL
AEKLRQOGCWVTAIRPTVPAGTARLRLITLAAHEWQDIDRLLEVLHGNG"

RBS
3030. .3045
/standard_name="BIOB RBS"
3043. .3753
/gene="BIOB"
3043. .3753
/gene="BIOB"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="PTB SYNTHASE"
/protein_id="CAA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYEYTGDTFVGKTVASCALLQAAGAKAGRTAGYKPVASGS
KTPGGLRNSDALALORNSSLQDLYATVNPYTFEAPSPHISAOEGRPIESLVMASG
LRALFOQADWVIVGAGAGWETPLSDFTFADWYQEOPLVILVGVKGLGINHAMLTA
QVIOHAGITLTAQWVANDVTPPGKRHAETMTLITRMIPAPLIGELPMLAENPENAAAGK
YINLAFVASTLGTFTSRL"

BASE COUNT 1318 a 1552 c 1695 g 1307 t

ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tcgtcaggtcaggtcagcagcttg 25
|||||
Db 227 TCCTCAGGTGTCAGGTGTCAGCAGCTTG 251

RESULT 8
A93674 5872 bp DNA PAT 22-JAN-2000
LOCUS A93674
DEFINITION Sequence 1 from Patent EP0798384.
ACCESSION A93674
VERSION A93674.1 GI:6741862
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O. and Brass, J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 1 01-OCT-1997;
LONZA AG (CH)
location/Qualifiers
1. .5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A-15/9"
1. .96
/function="PROMOTER PRAC"
/evidence=experimental
23. .28
/evidence=experimental
45. .49
/standard_name="PROMOTER PRAC"
105. .119
/standard_name="BIOB RBS NO. 9"
/evidence=experimental
117. .1157
/gene="BIOB"
117. .1157
/gene="BIOB"
/codon_start=1

```

/transl_table=11
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHPRWTLISQVTELEFPEKPLDLLFEQVYHROHFPPOVOST
LLSTKTCACPDCKCYCOSSRYKGTGLERLEMEVEQVLESARKAKAGSTRFCGAM
KNHFRDPTLEQWVOVKAMGLEACMTLGLTSESQAKRLANAGLDYNNHLDPSFE
VGNITTRTYOERDLEKVDAGIKVCSGIVGLGEVVDKAGLLQNLNLPPTPES
VPIIMLYKVKGTPLADDVDADFDPIRTIAVARIMPTSYVRLSAGREONEOIAC
FMAGANSIFYGCKLITTPNPEDEKDLQFLRKLINPQQTIVLADNNOQRLQALMT
PDTEVYMAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVNRKQATAAFGRAAHYEQHADLROSADALLAMPORKYT
HVLADGGPGMMSHMRERHQAOTALDSPMLVOAKQKADHYTLADIESIPLATA
TFDLANSILAVOMCGNLTALRELYRVKGVAFITLVQSLPELHQAQVADPRP
HANRELPDEIBOSLVGHYOHIIQPTITLMDALSMSRLKIGITHLAEGHDPRL
TRSOLORLQALAMPQOGRPLYTLGLVIARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONANONATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLAPDRHIIMHPYISMTSPYPVYVASEGCELLISGR
LVQSSWMAIHGYNHNPOLNMAKSQIDAMRSIVMGTTTHAPALBCKRIAMTPOP
LBCVPLADSSVAVEYAKMAALOYQAKDEARORLTFRNQYHGDITFGAMSCDDPN
MSLAKGYLPENLEAPQSRMDGEWDERDVGAFARLMAAHEHIAVITPEIYVAG
GMRYHPEWLKRIKICDREGILLADEIATGFGRTGLFACENAIADITLIGKAL
TGCTMTLSATLITREVAETISNGACGFMHGTFMGNPLACAAANASLILBSGMOO
QVADI EVOLEQOLAPARADENYADVRLGAGIVETTHFVNMAALQKFFVEGVNIRP
FGKLITMPTYILLPQOLRLTAANRAVODEFTFO"
5088..5100
/standard_name="ORF1 RBS"
5098..5574
/gene="ORF1"
5098..5574
/gene="ORF1"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDQDKLPHRHFVNGCYDGNISPHILAMDPVAGTK
SFVVCYDPAATPGSGMMHWVNNLPADTRVLPGFGSGILVAMPDGVLOTRTDFKGTG
TGAAPKGETHRYLITVHALDIRLIDVDEGASGAMVFNHFHSLASASITAMS"
5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
1552 C 1695 G 1307 T

```

```

ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 tgcgcagtgacagtcagcagcttg 25
Db 227 TCGTCAGCTGCAGCTGCAGCAGCTTG 251

RESULT 9
A93679 5872 bp DNA PAT 22-JAN-2000
LOCUS Sequence 6 from Patent EP0798384.
DEFINITION A93679
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O. and Brass,J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)
FEATURES
location/Qualifiers
1..5872
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONANONATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSMQEKTNALDARRADALRRRYPVACAGRMVLVADDOYLN
SSNLYGISHHPQILIRAMQOCEAFGIGSGSGHVSQY SVHQALEELAEWLGYSHA
LLFISGFANQAVTIAAMAKEDRIIAORLISASILEAASLSPSOLRRFANDVTHLAR
LIASPCQOQWVTEGVFSMDGSAPIAEIQOYQOHNGLMWDADAGTGVIEGSG
SCHWQKVPBELLVYTFEGKFGVSGAVALCSSTVADYLLQPARHLIYSTSMPPAOAL
RASLAVIRSDGDAERKIALILTRFAGVODLPTLADSCSALIOPLIVGDNRAIOL
AEXLRQGCWWTATIRPPTVPAGTARLRLTLTAHEMDIDRLLEVLHGNC"
3030..3045
/standard_name="BIOD RBS"
3043..3753
/gene="BIOD"
3043..3753
/gene="BIOD"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRTFVTGTDTIEVGKTVASCALLQAAKAGYRTAGYKPVASG
SKTPESGRNSDALQDNSSLOLDYATVNTYFAETSPHII SAQGRPIESLVMASG
LRALQOQWVILVAGAGCWPTPLSDFTFADWVTOQLPVLIVGVKLCGINHAMLTA
OVIOHGLITLAVANDVTPPGKRRAEYMTTLTRMIPAPLIGRIPLAENPEAAATGK

```

BASE COUNT 1318 a 1552 c 1695 g 1307 t


```

repeat_region
/db_xref="taxon:83333"
<1..130
/note="REP (repetitive extragenic palindromic) element;
contains 4 REP sequences"
complement(147..1430)
gene
/note="ybhc"
/note="b0772"
complement(147..1430)
CDS
/note="orf; Not classified"
/note="f427; 98 pct identical to fragment YBHC_ECOLI
SW:P46130 (300 aa) but contains 127 additional C-terminal
residues"
/codon_start=1
/transl_table=1
/product="putative pectinesterase"
/protein_id="AAC73859.1"
/db_xref="GI:1786989"
/translation="MNTFSVRLALAFGLTLAGSSPPDQPSDQAPGSSRPI
LSKKAQNFPAOHVPASTLPGAAMNPSPITTEPAQDPVVGAGTGTHTTTQAAVD
AATIKRTNKROYIAVMPGEYOGTYVPAPAGITLTGTGKPIDVIGLSLDGMSPA
DMRHDPNPKRTMPKRPAMYMDSCSRSDSGVLSAFAVMSQNNGLQNLQNTENT
LGSDVAGNHPAVALRTIDQDYOINNVNLTGRQNTPEPVNSGVQNNLETRNPRLVT
NSYIEGDVIDVSGRAVPDNTFEFRVNSRTOQEAFFAPATLSNIYGEFLAVNSFDV
AFQDGYAQLGRSLDVAQNTNGVYIRDSAINSEGTAKPMADAVISNRFAGNTGSVD
DNDEIQNLNDTVNRMWEYNRMVSGSKVAEAKK"
complement(1507..1535)
promoter
/note="factor Sigma70; predicted +1 start at 806574"
complement(1582..2058)
gene
/note="ybhc"
/note="b0773"
complement(1582..2058)
CDS
/note="orf; Unknown"
/note="f158; 99 pct identical to YBHC_ECOLI SW: P12994"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAC73860.1"
/db_xref="GI:1786990"
/translation="MKLISNDLRDCKLPHRIVFNGMGYDGNISPHLAWDDVPAGTK
SPVYTCYDDAPRTGSGWIRWVYNLPPADTRVLPQSGSLVAMPDGVLTQTRDFTGKG
YDGAAPKGETHYIETVHADIERIDVDEGASGAMVGFNVHSHIASASTAMS"
complement(2108..2124)
protein_bind
/note="central position to predicted promoter:85..5"
/bound_moiety="Rhas predicted site"
complement(2117..3406)
gene
/note="bioA"
/note="b0774"
complement(2117..3406)
CDS
complement(2117..3406)
/note="bioA"
/EC_number="2.6.1.62"
/function="enzyme; Biosynthesis of cofactors, carriers;
Biotin"
/note="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
/codon_start=1
/transl_table=1
/product="7,8-diaminopelargonic acid synthetase"
/protein_id="AAC73861.1"
/db_xref="GI:1786991"
/translation="MTTDDIAFDQDHIMHPTSMSPILPVVVSABCECILSDGRR
LVDGMSMAALHGINHPQLNAAKMSQIDANSHVFGGITAPALITCRKLVAMTPOP
LECYFLADSGVAEVAAMKALQYWAQGEARQRLITRNQYHGRTFGAMVCDPDS
MHSIMKGYLPENLEFAPASQMDGEMDEIRADWGFARLMAARHETIAAVITIEIVOGAK
GMRWYHPMLKRIKICDREGILADIADGFGTGLFACENAEETIAAVITIEIVOGAK
TGGWTLSATLTREVAETISNGEAGCFMGPTTGMGACANASTALIESGDMCO
OVADIEVOLRDLAPADAEMAVDRAVYGAIGVETTPRVMAALQRFYVQGWITP
FGKLIYMPYIILPQOLQRLTAANRAVQDETFFCQ"
complement(2193..2221)
promoter
/note="bioA"
/note="factor Sigma70; predicted +1 start at 807260"
3411..3450
/note="central position to bioB promoter: -20"

```

```

protein_bind
/bound_moiety="BioB documented site"
3411..3450
/note="central position to predicted promoter: -20; genetic
evidence for the site"
protein_bind
/bound_moiety="BioB predicted site"
complement(3411..3450)
/note="central position to predicted promoter:50"
/bound_moiety="BioB predicted site"
complement(3411..3450)
/note="central position to bioA promoter:50"
3413..3441
/bound_moiety="BioB documented site"
808525"
/note="factor Sigma70; promoter bioB; documented +1 at
complement(3447..3473)
/note="factor Sigma70; promoter bioA; documented +1 at
808515"
gene
3493..4533
/note="bioB"
/note="b0775"
3493..4533
/gene="bioB"
/EC_number="2.8.1.-"
/function="enzyme; Biosynthesis of cofactors, carriers;
Biotin"
/note="o346; 99 pct identical to BIOB_ECOLI SW: P12996"
/codon_start=1
/transl_table=1
/product="biotin synthesis, sulfur insertion?"
/protein_id="AAC73862.1"
/db_xref="GI:1786992"
/note="bioA"
4530..5684
/gene="bioA"
/note="b0776"
4530..5684
/gene="bioA"
/EC_number="2.3.1.47"
/function="enzyme; Biosynthesis of cofactors, carriers;
Biotin"
/note="o384; 100 pct identical to BIOF_ECOLI SW: P12998"
/codon_start=1
/transl_table=1
/product="8-amino-7-oxononanoate synthase"
/protein_id="AAC73863.1"
/db_xref="GI:1786993"
/translation="MSWQEKINAALDARRAADALRRRYPAQAGHMLVADDROYLNF
SSNDYGLSHHPQIIRAMQGAQGFQIGSGSGHVSQYSVHQAALFEELAEMLGYSRA
LFTISGPAQAOVIAAMAKEDRIADRLSHSLTAAASISQSLRFAHNDYTHAR
LLASPCGAQOVVTEGVFSMDGSDAPLATIGQVTOOHQMLWVDADHGCYVGEORG
SCMLQKKPELIVVTFGKGVSQAVALCSSYADVLILDFARHLITSTMPRAQOAL

```

```

Query Match 100.0%; Score 25; DB 1; Length 11022;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 tgcgcagtgaggtcagcacgtg 25
Db 3603 TCGTCAGGTGACGTCAGCACGTTG 3627

```

```

RESULT 13
SMABIO 7215 bp DNA BCT 04-FEB-1999
LOCUS S. marcescens biotin operon, complete sequence.
ACCESSION D17468
VERSION D17468.1 GI:402530

```

KEYWORDS
7,8-diamino-pelargonic acid aminotransferase;
7-keto-8-aminopelargonic acid synthetase; DAPA aminotransferase;
DTP synthetase; KAPA synthetase; biotin operon; biotin synthetase;
dethiobiotin synthetase.

SOURCE
Serratia marcescens (strain: Sr41) DNA.

ORGANISM
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Serratia.

REFERENCE
1 (bases 1 to 7215)
Sakurai, N.

AUTHORS
Direct Submission

JOURNAL
Submitted (25-AUG-1993) to the DDBJ/EMBL/GenBank databases. Naoki Sakurai, Tanabe Seiyaku Co., Ltd., Res Lab of Applied Biochemistry; 2-50, Kawagishi-2-chome, Toda-shi, Saitama 335, Japan (E-mail: GFC01101@niftyserve.or.jp, nsakurai@ddj.nig.ac.jp, Tel: 048-433-2545, Fax: 048-433-2540)

REFERENCE
2 (bases 1 to 7215)
Sakurai, N., Imai, Y., Akatsuka, H., Kawai, E., Komatsubara, S. and Tosa, T.

TITLE
Complete nucleotide sequence of biotin operon of Serratia marcescens

JOURNAL
Unpublished (1993)

COMMENT
Submitted (25-Aug-1993) to DDBJ by:
Naoki Sakurai
Res. Lab. of Applied Biochemistry
Tanabe Seiyaku Co., Ltd.
2-50 Kawagishi-2-chome
Toda, Saitama 335
Japan
Phone: 048-433-2545
Email: nsakurai@ddj.nig.ac.jp
Fax: 048-433-2540.

FEATURES
source
1. 7215
Location/Qualifiers
/organism="Serratia marcescens"
/strain="Sr41"
/db_xref="taxon:615"
complement(1353..2630)
/gene="bioc"
complement(1353..2630)
/gene="biocA"
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/product="7,8-diamino-pelargonic acid aminotransferase
(DAPA aminotransferase)"
/protein_id="BA04284.1"
/db_xref="GI:402531"
/translation="MSVTASDIAFDQRHITMPTSMRPLPCYPIESASGVLEOLADG
RSIVDGRSSWMAAIIHCYNHPRILMOASQCEKSHVPGITRPAATISICRLVEMTP
EALOCYFLADSGVAEVSLSKALQYWAQGERKQITLIRHGYHGDITGAMSVCPD
NSMHSLYQGLAPHLFATAPQCRFDEWREDEIAPRALIEHGAEEAAVILPEVVG
AGGMRIYHPYLRKRVDRVVAIKLLIADEIATGCRGKFLFACSEAQVADPILCLG
KALTGMYTSLATLTRHVAETISNGAAGFMHGPTFMGNPLACAVADSLALAEHR
WQOVSAIETQLRELLPLGRCRWPHVRLGAGVEMKEPDAVELQGEVERGVW
IRPEKTIYIMPTIIEADQLSLTAADVAAAR"

gene
2716..3756
/gene="bioc"
2716..3756
/gene="biocB"
/codon_start=1
/transl_table=11
/product="biotin synthetase"
/protein_id="BA04285.1"
/db_xref="GI:402532"
/translation="MMADRIMHTVGAQALFDKPLLELLFEAQTVHRHPPQVOVS
TULSTKGACDECKYCPQSPRYKTGLSESLRMQVLEVSARKAKANGSTRCMAA
WKNRHEMDMYLOMGVCKAMGEMTCMTGITDGAERLAEAGLDYNYHNDTSP
PYGSIITTRSYQERLDTLDKYRDAGIVCGSGIVGCEYTRDRAGLLVQANLKPPE
SVPLINMLVKVKGPLAPNDVDVDPDFTRTIAVARIMPPSSIVRLSAGRDMDQTM
CFMAGANSIFYGCKLITTPNPEEDKDLQRLKRLGLNPOQTATEHGNDQOVLAKOLL
NADTAFFYNAAP"
3756..4904
gene

CDS
/gene="bioc"
3756..4904
/gene="bioc"
/codon_start=1
/transl_table=11
/product="7-keto-8-aminopelargonic acid synthetase (KAPA
synthetase)"
/protein_id="BA04286.1"
/db_xref="GI:402533"
/translation="MSWQRTIEQALERRILMAAYRRRTTEGGNGRQIRIGDRLTYLNF
SGNDYGLISDARVIAAMQDQARYGVNGSGSHYTFSAHQALDEQLAAMLYPRAL
LFISYANQAVLAIMQGDRIIADRLSHASLLEAAGAVGRAPVPAQSTAGLARI
LAKPCDQRLAATEGLGSMGDAPLAEHLRLTRAGAAWLVDAHGIYRGQGRKS
CMQGVPELLVATFAGFAGVGAAYCDEATAEVLLQFRRHIIYSTAMPADACALO
AALARIQEDDRLARLQDNIIRRFQGAAPLALITLDSDTALQPLVAGDNRALDLATR
LREGGLVSAIRPPTVPFGARLRLLTAAHQSDIDRLLEVLNDVSO"
4888..5655
/gene="bioc"
4888..5655
/gene="bioc"
/function="synthesis of pimeloyl-CoA"
/codon_start=1
/transl_table=11
/product="the product of bioc"
/protein_id="BA04287.1"
/db_xref="GI:402534"
/translation="MTSANDVFNKQAVASFRAGSYDAALALORDVGERLLGMSS
HPGRLDAGCGTGYSRMREKGRVYALDARGMIDVARORQAHHYLLQDEQV
LPDAAMDIFSSLVQWCSIDLPALAEILRYTRRGVYLPSTLAGSLQELGDMQV
DGERVNAFLPDTQIRTAQAVRHELVTELTATVDPVMTLMSLKGIGATHLQGRE
GGLMSGRGLAALQAAVPCROGQFPLSYHLAVGYIRE"
5648..6331
/gene="bioc"
5648..6331
/gene="bioc"
/function="conversion from 7,8-diaminopelargonic acid into
dethiobiotin"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase(DTB synthetase)"
/protein_id="BA04288.1"
/db_xref="GI:912453"
/translation="MSKRWFTYGTDEYGVKSVASSALLQAAANRAGYPSAGKPYASGS
EWTAEGLRNGDALQANSVGLADYDEVNAPYVFEPPSPHIVSADDEGPIDAARLSDG
LRLRQADWDVTEGACGMFTPLSAEYTFADWVROELPVLLVYGIKIGCYNHVLTA
QAVQOAGLTIAGMINDVGAAPGRRHQYLTATLRMLPRCMASRTRRPRRPLQ
YIDISLQAO"

BASE COUNT
1347 a 2225 c 2367 g 1276 t

ORIGIN

Query Match
Best Local Similarity 100.0%; Score 21; DB 3; Length 7215;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
2830 CGTCAGGTCGAGTCACGACG 2850

RESULT 14
AE005258 13501 bp DNA BCT 21-MAR-2001
LOCUS
DEFINITION
Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
of 155.
ACCESSION
AE005258 AE005174
VERSION
AE005258.1 GI:12513751
KEYWORDS
SOURCE
Escherichia coli O157:H7 EDL933.
ORGANISM
Escherichia coli O157:H7 EDL933
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE
1 (bases 1 to 13501)

AUTHORS	Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Postel,G., Hackett,J.F., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobleck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
TITLE	Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
JOURNAL	Nature 409 (6819), 529-533 (2001)
MEDLINE	21074935
PubMed	11206551
REFERENCE	2 (bases 1 to 13501) Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Postel,G., Hackett,J.F., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobleck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
FEATURES	Direct Submmission Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
source	Location/Qualifiers 1..13501 /organism="Escherichia coli O157:H7 EDL933" /strain="EDL933" /serotype="O157:H7" /db_xref="taxon:155864" /note="enterohemorrhagic" /note="O-island #36; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655; Cryptic prophage CP-933K; Includes one copy of the 13 bp direct repeat that flanks the prophage" 66..665 /gene="lomk" /note="Z0981"
gene	/function="putative membrane; Other or unknown (Phage or Prophage Related)" /note="Residues 1 to 199 of 199 are 72.86 pct identical to residues 1 to 199 of 199 from Genpept 118 : g117532789 gb AAFG3231.1 AFI51091_2 (AFI51091) lom [prophage P-EibA]" /codon_start=1 /transl_table=11 /product="putative outer membrane protein of prophage CP-933K" /protein_id="AAG55137.1" /db_xref="GI:12513752"
CDS	/translation="MRKCAIIISNAICIAVSGVPWASEHSTLSAGTILHASTDAPG PSYRNENFSATIMAGAVSKVSFSDYPERVDNKRRTHDVLITGSDARYSNMISLAWGAGVENFTESVAADVAYEXSGDMWRIDFIVGVYKF" 719..2041 /gene="Z0982" 719..2041 /gene="Z0982" /function="putative structure; Structural component (Phage or Prophage Related)" /note="Residues 164 to 440 of 440 are 68.79 pct identical to residues 381 to 645 of 645 from Genpept 118 : g114585437 gb AAD25465.1 AFI25520.59 (AFI25520) putative tail fiber protein [Bacteriophage 933w]" /codon_start=1 /transl_table=11 /product="putative tail component of prophage CP-933K" /protein_id="AAG55138.1" /db_xref="GI:12513753"
gene	/translation="MGMAVOISGLVKDGAKPIQNTIOIKARRSTYYVVYVTFASNPDAGRSMDEVXGOYSVTLLVGGPPSHAGTISYEQQPTLNDPLGAHTEDAR PEALIRPEOMWEPAARIHAEFRKKRAGAEITSARNAGISAKASAANADTSIEDAR SARQAABSAFAKSEBSASSSEAOKASESLQSQTADAEISKTAESAAGAARDATAITTEKREAESSESQSAEQSRIAAEAVNRIPVVGPPGKGEPGPGAGCGPGDGE
misc_feature	<1..7576 /note="O-island #36; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655; Cryptic prophage CP-933K; Includes one copy of the 13 bp direct repeat that flanks the prophage" 66..665 /gene="lomk" /note="Z0981"
gene	/function="putative membrane; Other or unknown (Phage or Prophage Related)" /note="Residues 1 to 199 of 199 are 72.86 pct identical to residues 1 to 199 of 199 from Genpept 118 : g117532789 gb AAFG3231.1 AFI51091_2 (AFI51091) lom [prophage P-EibA]" /codon_start=1 /transl_table=11 /product="unknown protein encoded by prophage CP-933K" /protein_id="AAG55140.1" /db_xref="GI:12513755"
CDS	/translation="MLSPTRPTPHNSVNIVOSSPCOTVSPAGKEYLKVIDEKPTILF OMFEENPRRYKKDEVIPTNKOPHYLDINVMNAIESDRMIGIFVGDGPENVKTAPS KLDERFENMUIREDYDFSWYDKRLSDIHDIIDTCERLRLETKRKDETLNLLEKEIR ELISKODSLIMKYAKRRHNHFDFRRNALALKAGEITRCYTNTKNIGTISFGECITYLD MDMLITGKLGTITVAPDGISMHVDRNDSVNTENSAILVNSNHPALLEGLSFPMHSKYD AHYPYDGLGKGVCKVFNFPTLHNYNHFCDFIEFWHPNI IMNTSOYCCSSW" 3530..4522 /gene="Z0986" 3530..4522 /gene="Z0986" /function="orf; Other or unknown (Phage or Prophage Related)" /note="No significant matches" /codon_start=1 /transl_table=11 /product="unknown protein encoded by prophage CP-933K" /protein_id="AAG55141.1" /db_xref="GI:12513756"
gene	/translation="MKLIPSIQSNFNESAPAGYCAPAPIAPRNAENAVALDYLDIGKRILP SAADLSNVYESIVRAVHDSRSRLIDQFTVMTINTVLDLASRQTPRDVAHSVGINHEK VHIQCIKYRNEYELNEESSVYKIDIDISLTICNELYEVDCOEPIPTICEAGENDNEPY VSFSVADPTSYEMPDSMOEGLIHETIIHHVYGSSDPGSDSNIELGPEITELIARRAOELG WSVDFKGYAPEPEREHLRLRNALROAMRHEENRAREFEELTGITISRVEASPDT EYSAVSNIGYGFILQHFPGLAINDNLNQDANOQLYHGAPYIFTFDYVKNHQQ" 5350..6231 /gene="Z0989" 5350..6231 /gene="Z0989" /function="orf; Other or unknown (Phage or Prophage Related)" /note="No significant matches" /codon_start=1 /transl_table=11 /product="unknown protein encoded by prophage CP-933K" /protein_id="AAG55142.1" /db_xref="GI:12513757"
CDS	/translation="MISPVSVNIGCSWNSTITWTITSENBNUTISVUDNAVHEDNNKOV

KVGNRTYRVATDNKFCVTRSHSGCTNLHLRLGMPKGEISRKLEVMNLSPYSAM
 ERGLVHSNRPLDLPVDPVAPPELPSVDYVNRSLPGVIGKGNNAVYEDAEATYVUKM
 FTTSSQNEDEVEKFCNQYGGASAEITGNNGDIICIRDRKINGESLNLISLIPAO
 AEHAIYDMDFLEOKGILFVDTETNVLVYDRAKNEFIDISYVWSDRSSESIOIMO
 SYHGKODLISVLSKI"

gene
 /gene="20990"
 /length="7160"
 /function="orf; Other or unknown (Phage or Prophage Related)"
 /note="No significant matches"

codon_start=1
 /transl_table=11
 /product="unknown protein encoded by prophage CP-933K"
 /protein_id="AAG55143.1"
 /db_xref="GI:12513758"

/translation="MRPMSLNVLHSSSSSSMDPDIESLKASSVQIMKNNPOLRF
 OGTDNITVOIEAALDKISGTGTGKVLNLAISISRLKSEYVILHNSRLGVMARD
 IDAENHRTGSDPHCHLNAVERPCGSGISVDFHATVPHELLHFHNLNGRLKVES
 SRAESQKSPILLEEARIVGIGAFSEEVISENKFHEIGMPRTSPYDPSALIHDDNT
 VSLGFQVRLHPL"

gene
 /gene="yphb"
 /note="20992"
 complement(7668..8144)

/function="orf; Unknown function"
 /note="Residues 1 to 158 of 158 are 99.36 pct identical to
 residues 1 to 158 of 158 from Escherichia coli K-12 Strain
 MG1655: H0773"

gene
 /gene="yphb"
 /note="20993"
 complement(8203..9492)

gene
 /gene="yphb"
 /note="20993"
 complement(8203..9492)

Query Match 84.0% Score 21: DB 1: Length 13501:
 Best Local Similarity 100.0%: Pred. No. 0.19:
 Matches 21: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 1 tcgtcaggtcaggtcagcac 21
 |||||||||||||||||||||
 Db 9689 TCGTCAGGTGCGAGCTGACGAC 9709

RESULT 15

AP002553 297816 bp DNA BCT 07-MAR-2001
 LOCUS Escherichia coli O157:H7 DNA, complete genome, section 4/20.
 DEFINITION AP002553 BA000007
 ACCESSION AP002553.1 GI:13360211
 VERSION
 KEYWORDS

SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)

ORGANISM Escherichia coli O157:H7
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE 1 (sites)
 Makioka, K., Yokoyama, K., Kubota, Y., Yutsudo, C.H., Kimura, S.,
 Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T.,
 Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
 Sasaki, C., and Shinagawa, H.

AUTHORS Complete nucleotide sequence of the prophage VT2-Sakai carrying the
 verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
 derived from the Sakai outbreak

JOURNAL Genes Genet. Syst. 74 (5), 227-239 (1999)
 MEDLINE 20198780
 REFERENCE 2 (sites)
 Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M.,
 Kurokawa, K., Yasunaga, T., Yokoyama, K., Makioka, K., Shinagawa, H. and
 Hayashi, T.

Comparative analysis of the whole set of RNA operons between an
 enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
 Escherichia coli K-12 strain MG1655
 Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
 20557356
 3 (sites)
 Yokoyama, K., Makioka, K., Kubota, Y., Watanabe, M., Kimura, S.,
 Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
 Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasaki, C. and
 Shinagawa, H.

Complete nucleotide sequence of the prophage VT1-Sakai carrying the
 Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
 O157:H7 strain derived from the Sakai outbreak
 Gene 258 (1-2), 127-139 (2000)
 20564182
 4 (sites)
 Hayashi, T., Makioka, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
 Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T.,
 Tanaka, M., Tobe, T., Taki, H., Takami, H., Honda, T., Sasaki, C.,
 Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and
 Shinagawa, H.

Complete genome sequence of enterohemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12
 DNA Res. 8 (1), 11-22 (2001)
 21156231
 5 (bases 1 to 297816)
 Ohnishi, M., Kurokawa, K., Makioka, K., Yasunaga, T., Shinagawa, H. and
 Hayashi, T.

Direct Submission
 Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
 Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871,
 Japan (E-mail: ken@gen-info.osaka-u.ac.jp),
 URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,
 Fax: 81-6-6879-2047

COMMENT Location/Qualifiers
 FEATURES
 source

gene
 /organism="Escherichia coli O157:H7"
 /strain="O157:H7"
 /sub_strain="RIMD 0509952"
 /db_xref="taxon:83334"
 79..1245
 /gene="ECS0753"
 79..1245
 /note="similar to SUCC_ECOLI g11786948 percent identity
 100 in 388 aa (Conserved in E.coli K-12)"

gene
 /organism="Escherichia coli O157:H7"
 /strain="O157:H7"
 /sub_strain="RIMD 0509952"
 /db_xref="taxon:83334"
 79..1245
 /gene="ECS0753"
 79..1245
 /note="similar to SUCC_ECOLI g11786948 percent identity
 100 in 388 aa (Conserved in E.coli K-12)"

gene
 /organism="Escherichia coli O157:H7"
 /strain="O157:H7"
 /sub_strain="RIMD 0509952"
 /db_xref="taxon:83334"
 79..1245
 /gene="ECS0753"
 79..1245
 /note="similar to SUCC_ECOLI g11786948 percent identity
 100 in 388 aa (Conserved in E.coli K-12)"

gene
 /organism="Escherichia coli O157:H7"
 /strain="O157:H7"
 /sub_strain="RIMD 0509952"
 /db_xref="taxon:83334"
 79..1245
 /gene="ECS0753"
 79..1245
 /note="similar to SUCC_ECOLI g11786948 percent identity
 100 in 388 aa (Conserved in E.coli K-12)"

gene
 /organism="Escherichia coli O157:H7"
 /strain="O157:H7"
 /sub_strain="RIMD 0509952"
 /db_xref="taxon:83334"
 79..1245
 /gene="ECS0753"
 79..1245
 /note="similar to SUCC_ECOLI g11786948 percent identity
 100 in 388 aa (Conserved in E.coli K-12)"

gene
 /organism="Escherichia coli O157:H7"
 /strain="O157:H7"
 /sub_strain="RIMD 0509952"
 /db_xref="taxon:83334"
 79..1245
 /gene="ECS0753"
 79..1245
 /note="similar to SUCC_ECOLI g11786948 percent identity
 100 in 388 aa (Conserved in E.coli K-12)"

gene
 /organism="Escherichia coli O157:H7"
 /strain="O157:H7"
 /sub_strain="RIMD 0509952"
 /db_xref="taxon:83334"
 79..1245
 /gene="ECS0753"
 79..1245
 /note="similar to SUCC_ECOLI g11786948 percent identity
 100 in 388 aa (Conserved in E.coli K-12)"

/transl_table=11
/evidence=not_experimental
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/translation="MSLIDKNTKVICOGFTGSGQTFHSECAINVGTKMGVYPRGK
GTHHGLIPFNTVREPAVATGATASVITYRPFKDSILEIDAGIKLITITTEGIP
LDMITVKKKIDKDEAVRMIGPNCVITPGECKIGIQPHHKPKGVIVSRSGTLTVE
AVKQTDGFGQSTCVGIGDIPGSMNFDILEMFKEDPQTEALVMIGELGSGMEEA
AAVTKEHVTKPVGVYIAGVTPAKGKRMGAGAIATAGCKTADKFALEAAGVKTWS
LADIGELAKTVLK"
2781..3686
/gene="ECs0755"
2781..3686
/note="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
Escherichia coli g11417043|sp|P32064|GCVA_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"
/translation="MRGKIPKTELIVFEVVAHRESTRABEALATQSAVPRQVSAI
EERLPLFSSKRRIFLNDTKTYTIVETLAKLERTDNTIMTQPTVOVELAVN
PTESTHMLFMLEFTHLHPDILVNIHSLANNDQFNREYDAVIMREFCAPAAVEY
LFEELIIPVCSGSLAMSDOKLSVAELITELPLIHQSTRTGMEWELGSGVSPV
NGRFDLMLIAVRSNLGVALPRFAIQHDDSDGMVPCDVPRTGRTFMWGE
EKSDSPHQOFREWLAKSVVPM"
complement(3720..4322)
/gene="ECs0756"
complement(3720..4322)
/note="probable cob(I)alamin adenosyltransferase, similar
to cob(I)alamin adenosyltransferases (corrinoid
adenosyltransferases) e.g. [Escherichia coli]
g11151481|sp|P130401|BTUR_ECOLI percent identity 67 in 200
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative cob(I)alamin adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/translation="MEARISTERNVOROKLIKPOVTRVAAALEKKGLIVPTGNGK
KSTAFRTVTRAVGHGKTVGAOYIKGOMNGEYNNLOPLGVGFHMGTFWETQNR
QADIDAKREVMSKRLADKRYDLVLDLETYMLAHYLDTEEVATSIQNRPAQGSV
IVTGRGCHSOILKMDTVEIRPVKHAHPNGIOAQPGIDW"
complement(4332..5984)
/gene="ECs0757"
complement(4332..5984)
/note="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
(fumarate) - Escherichia coli
g11205981|sp|P00923|FUMA_ECOLI percent identity 68 in 545
aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"
/translation="MSKPRIMQBPLOKQDTEYLLISDQHTVTELDGEVYKTAPE
ATILLARQAFYASFFLRSALHQAQVASTINDQASSNDYVALOLLRAAEVSAKCVLP
NCODGTATVASKGQQLWTGNDALSKGTYSTFOENMLFSGNAPLDVTEVNTQ
TNPAQIDISAAVGEYHPLCVNKGSGSANKALYQETKSLQPEKTLAFLEKKSL
GTAACPGYHIAVGVGLSADQTLAKLASTKYNNLPDPTSGMNGQARDDILEKVL
EASQOFGIAAGGKGFADHDIVIRLPRHSGSCPTAMALSCADNCAIKKIGTIL
EKLHNPGYITPASLREBNHQAQVOLDLNRPLRDVMDLRLPVGTRVSLSGPIVAVR

DIANAKIKARLDSGEPMPEYLKHHIVYVAGAPTEBNMAGSLGPTGGRMGYIDTF
OAGGSLVMSLKGNSQOYVADACHKRGNGNLSIGSAAALLAQEVYKSLRLEYELG
MEAVMMVEYENLPATFLVDDCKNNFSSQEQHRCASCPAGH"
complement(6092..7372)
/gene="ECs0758"
complement(6092..7372)
/note="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. [Bacillus stearothermophilus]
g1121467|sp|P24943|GLUT_BACT percent identity 38 in 416
aa, also similar to C4-dicarboxylate transport"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translation="MKKISLTMTMLLAVLGMITGVNLTNMTASPEAKVAQETSIPT
TIFLRILKMIAPLVVSTLVGIAKMGAKALGRIFSTLFLICASLISLALGLITV
NFFMPQGTGTFVAGHGAETGVVAAPFTLKVFISHAPTSIVDAMAHNEIQYVESI
FLGCSLTALGEGKSAIVHALLSLAHAMLIKGVYMLFAPLPAFASLALMERGLAV
VSAGTFMGEFFETMLLWYLIGLAIVYVGPCIRRTVALSERPALATTSSRAAP
GLEKLEQGVSPKIASVYLPVIGYFNLVGSMAKCSFATVTAQCNTHLSITGQITM
LILMLTSGMAGVPRASVNVVIAATLQNFNPEAGLILMGVDPDLMGRSATVMSN
AAGAAMVSRBEHGFEGCGRCALKPNESNALP"
complement(7533..7850)
/gene="ECs0759"
complement(7533..7850)
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"
Query Match 84.0%; Score 21; DB 2; Length 297816;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tcgtcagtcacgagtcacac 21
|||||
Db 92421 TCGTCAAGTGCAGTCACGAC 92441
RESULT 16
AE001862/c DNA BCT 22-NOV-1999
LOCUS Delinococcus radiodurans R1 section 1 of 2 of the complete
DEFINITION chromosome 2.
ACCESSION AE001862 AE001825
VERSION AE001862.1 GI:6460468
KEYWORDS Delinococcus radiodurans.
SOURCE Delinococcus radiodurans.
ORGANISM Bacteria; Thermus/delinococcus group; Delinococcales; Delinococcus.
REFERENCE 1 (bases 1 to 213732)
White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,
Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,
Moffat, K.S., Qin, H., Jiang, L., Pamphile, M., Crosby, M., Shen, M.,
Vamathevan, J.J., Lam, P., McDonald, L., Uitterback, T., Zalewski, C.,
Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M. et al.
Genome sequence of the radioresistant bacterium Delinococcus
radiodurans R1
Science 286 (5444), 1571-1577 (1999)
2 (bases 1 to 213732)
White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,
Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,
Moffat, K.S., Qin, H., Jiang, L., Pamphile, M., Crosby, M., Shen, M.,
Vamathevan, J.J., Lam, P., McDonald, L., Uitterback, T., Zalewski, C.,
Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W.,

AAPSPQVEVRLLEGOLMOPGLGIPRPGPRSRVPELVTKRRPADIRLALYPPPPAL
 RELLEQRRPTGDFGSLRRTTMDLRAPGQARVAVHAEPAAGKIDALIVPDRIL
 VEINQENKSRDEIVEMISRGSDSVRLKQPISELSISMSVIRKTEGSHREBADA
 KTEBOIAEAMVETEKWLVHROFSLASOLSKSELSPEKAVIRKIDHGAALDIDV
 EBDITKANAPSCDRLEDAISLVYVNESSVILTRORVASCALITVAGSGLVLTIRGA
 PAVISERKWHNPGKCRREDMARIYAVAQATRAMDMRODSYLLSSSGKVTSP
 OHLVQYLATIAGTSGTVFSEKWOALSTLEAFNENSPIMNGATRSQISLIDPO
 AGOVASASIOIOMLEKLRVARRPASEATFNVEYVILAGDATALRHLNHLNENVE
 GIVPLSKPEKQAAQOFKIOAMKVLAISPEEDQTCMLIASIYHGAATNVEEA
 GKRQOPARHEMAOKAAVLLGSLSELSAIFKHOLKGTIORSTSPFGQPEESGLGE
 GTRKASLECEBEMASGIVSELPITLISIVNALKSSOHSICSMIVDPFGONPMWG
 SARQASFEELCHNYADRIQRFHETPTQELERKKNIEELAFDIDLEPVADSVAAV
 DASHLVRSLAHADERGLMLLEBALVGCATEDALDRLETSYGPORBDKQSGPL
 LSSKRRHFLGHSHGTNVEYNAGVWNTKONPATONAPRLDSDOKKIIISNPLFG
 RASGATVLSGSIAGLEGQSOLARATSMRKPTTGMAVKKRSLCLOIKILOVDLID
 TIKRSMHVFHCEFLPVAEGMPGEPSSRSRVSSELDLPDGCDEAGLLOLDVSL
 ROLRSGRLLDMARMYROGPDHMFSEFRREDVLAPLTKKHGNTYVVEKRAVE
 ELLESIDLEKSSCCGSLSPFRAGTLALFEORBDQTSRHILFOACRGVLAHOH
 KRRIODDLATRCOKITKKNKGVKQDPMPKLTTPYRPLIOVLOLSEOFKPKREIOL
 RSKLEVERERNELRLSDRLRISLSELTDEKNTGESLADGAGEWRLKRYARAV
 ESKELQTOYDALKKOMEVMEVMEARILRAEINGEVDDDSAGGELWRLKRYARAV
 DFKRLQOOLEDKMEVEQOSROLERLGLDQADSDSODALQOLKQKCOBLTELQ
 DKTLHGOOVNRHLEKKORRPSLSQAHETOREKLOREKLOREKMLIAEAFSL
 KOOMEKDIAGFTOKVYVLEAELODISOSKDEASLAKYKOLRODLEAVKQDEE
 ELDEBQSGIOMLEQALRLIEMEREMROTHSKEMESRDEVEAROSCKKIKOMEVO
 LEEYEDKOKALEREKRELSKSLSTSDQVNRDFSEKRLRDKRTKRLILDAOIML
 DLKNNAPSKRELAOLKNOLESEFTCAAVARKMEVEMEDLHIDIDIAKAKTAL
 EEOISLROREKNEIOMRLEDEQDMNLEAKKHAQAASRMAQMDLOAOIESNK
 EKOLOEKLOALSOVEFLQSVQVDSKVSROAKIRELETRLEFEKTOVKRIENIAS
 RIKETMEKLTREERDORAAAREKREKONKROLOKRTKEEMSLARKAEASRKIEL
 EMDLESELANOSIADIKLAFKRTIDLOALIEDMESENEDLINSSEDSVDSLE
 DRYDGVKSWLSKNGPSKAPSDSGSLKSSPSSHMKPLAPDPSPDDHDVDVDSISRF
 SHSYLSDSDTEALTELETSA

repeat_region
 /note="KE repeat"
 68..179
 misc_feature
 /gene="MysPDZ"
 /note="pdr domain"
 1272..3358
 misc_feature
 /gene="MysPDZ"
 /note="myosin head"
 1544..1565
 misc_binding
 /gene="MysPDZ"
 /bound_moiety="ATP/GTP binding site"
 3602..3665
 misc_feature
 /gene="MysPDZ"
 /note="IQ domain"
 3777..5865
 misc_feature
 /gene="MysPDZ"
 /note="coiled-coil"
 7279..7284
 polyA_signal
 polyA_site
 7303
 BASE COUNT 1768 a 1954 c 2232 g 1364 t
 ORIGIN
 Query Match 68.0%; Score 17; DB 94; Length 7318;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 cgtcagtcagtcag 18
 |||||
 Db 288 CGTCAGTGCAGTCAG 272
 RESULT 19
 ACC021767/c 69900 bp DNA HTG 13-JUL-2000
 LOCUS Mus musculus chromosome 11 clone RP23-47D19 map 11, LOW-PASS
 DEFINITION
 SEQUENCE SAMPLING.
 ACCESSION AC021767
 VERSION AC021767.2 GI:7144812
 KEYWORDS HTG; HTGS_PHASE0.

SOURCE house mouse.
 ORGANISM Mus musculus
 REFERENCE Eumariyola; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 69900)
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE Mus musculus chromosome 11, clone RP23-47D19
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 69900)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Beckery, R., Beda, F.,
 Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A.,
 Choe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
 DeRellano, K., Dewar, K., Domino, M., Doyle, M., Fenster, J.,
 Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
 Gardina, S., Grant, G., Hagos, B., Heald, A., Horton, L.,
 Howland, J., Johnson, R., Jones, C., Kam, L., Karatas, A., Klein, J.,
 Lander, T., Lehotzky, J., Levine, R., Lien, C., Liu, G., Locke, K.,
 Macdonald, P., Marquis, N., McEwan, P., McGuck, A., McKernan, K.,
 McPeckers, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J.,
 Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,
 Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
 Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J.,
 Tirrell, A., Vassiliou, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
 Zimer, A. and Zody, M.
 DIRECT SUBMISSION
 Submitted (20-JAN-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 3, 2000 this sequence version replaced gi:6721397.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L3201
 Center clone name: 47_D19
 * NOTE: This record contains 79 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1
 788 887: gap of 100 bp
 888 1651: contig of 764 bp in length
 1652 1751: gap of 100 bp
 1752 2546: contig of 795 bp in length
 2547 2646: gap of 100 bp
 2647 3438: contig of 792 bp in length
 3439 3538: gap of 100 bp
 3539 4327: contig of 789 bp in length
 4328 4427: gap of 100 bp
 4428 5234: contig of 807 bp in length
 5235 5334: gap of 100 bp
 5335 6128: contig of 795 bp in length
 6130 6229: gap of 100 bp
 6230 7040: contig of 811 bp in length
 7041 7140: gap of 100 bp
 7141 7931: contig of 791 bp in length
 7932 8031: gap of 100 bp
 8032 8810: contig of 779 bp in length
 8811 8910: gap of 100 bp

```

*      8911      9696: contig of 786 bp in length
*      9697      9796: gap of 100 bp
*      10597     10596: contig of 800 bp in length
*      10597     10696: gap of 100 bp
*      11494     11493: contig of 797 bp in length
*      11594     11593: gap of 100 bp
*      12400     12399: contig of 806 bp in length
*      12500     12499: gap of 100 bp
*      13297     13296: contig of 797 bp in length
*      13397     13396: gap of 100 bp
*      14153     14152: contig of 756 bp in length
*      14253     14252: gap of 100 bp
*      15035     15034: contig of 782 bp in length
*      15135     15134: gap of 100 bp
*      15923     15922: contig of 788 bp in length
*      16023     16022: gap of 100 bp
*      16796     16795: contig of 773 bp in length
*      16896     16895: gap of 100 bp
*      17699     17698: contig of 803 bp in length
*      17799     17798: gap of 100 bp
*      18569     18568: contig of 770 bp in length
*      18669     18668: gap of 100 bp
*      19458     19457: contig of 789 bp in length
*      19558     19557: gap of 100 bp
*      20343     20342: contig of 785 bp in length
*      20443     20442: gap of 100 bp
*      21240     21239: contig of 797 bp in length
*      21340     21339: gap of 100 bp
*      22123     22122: contig of 783 bp in length
*      22223     22222: gap of 100 bp
*      22996     22995: contig of 773 bp in length
*      23096     23095: gap of 100 bp
*      23837     23836: contig of 741 bp in length
*      23937     23936: gap of 100 bp
*      24721     24720: contig of 784 bp in length
*      24821     24820: gap of 100 bp
*      25591     25590: contig of 770 bp in length
*      25691     25690: gap of 100 bp
*      26481     26480: contig of 790 bp in length
*      26581     26580: gap of 100 bp
*      27369     27368: contig of 788 bp in length
*      27469     27468: gap of 100 bp
*      28259     28258: contig of 790 bp in length
*      28359     28358: gap of 100 bp
*      29135     29134: contig of 776 bp in length
*      29235     29234: gap of 100 bp
*      30017     30016: contig of 782 bp in length
*      30117     30116: gap of 100 bp
*      30898     30897: contig of 781 bp in length
*      30998     30997: gap of 100 bp
*      31771     31770: contig of 773 bp in length
*      31871     31870: gap of 100 bp
*      32663     32662: contig of 792 bp in length
*      32763     32762: gap of 100 bp
*      33559     33558: contig of 796 bp in length
*      33659     33658: gap of 100 bp
*      34449     34448: contig of 790 bp in length
*      34549     34548: gap of 100 bp
*      35347     35346: contig of 798 bp in length
*      35447     35446: gap of 100 bp
*      36252     36251: contig of 805 bp in length
*      36352     36351: gap of 100 bp
*      37137     37136: contig of 785 bp in length
*      37237     37236: gap of 100 bp
*      38021     38021: contig of 785 bp in length
*      38121     38121: gap of 100 bp
*      38122     38122: contig of 787 bp in length
*      38909     38908: gap of 100 bp
*      39009     39008: contig of 792 bp in length
*      39801     39800: gap of 100 bp
*      39901     40683: contig of 783 bp in length
*      40684     40783: gap of 100 bp
*      40784     41551: contig of 768 bp in length

```

```

*      41552     41651: gap of 100 bp
*      41652     42409: contig of 758 bp in length
*      42410     42509: gap of 100 bp
*      42510     43302: contig of 793 bp in length
*      43303     43402: gap of 100 bp
*      43403     44204: contig of 802 bp in length
*      44205     44304: gap of 100 bp
*      44305     45097: contig of 793 bp in length
*      45098     45197: gap of 100 bp
*      45198     45976: contig of 779 bp in length
*      45977     46076: gap of 100 bp
*      46077     46856: contig of 780 bp in length
*      46857     46956: gap of 100 bp
*      46957     47735: contig of 779 bp in length
*      47736     47835: gap of 100 bp
*      47836     48609: contig of 774 bp in length
*      48610     48709: gap of 100 bp
*      48710     49519: contig of 810 bp in length
*      49520     49619: gap of 100 bp
*      49620     50405: contig of 786 bp in length
*      50406     50505: gap of 100 bp
*      50506     51282: contig of 777 bp in length
*      51283     51382: gap of 100 bp
*      51383     52152: contig of 770 bp in length
*      52153     52252: gap of 100 bp
*      52253     53044: contig of 792 bp in length
*      53045     53144: gap of 100 bp
*      53145     53941: contig of 797 bp in length
*      53942     54041: gap of 100 bp
*      54042     54836: contig of 795 bp in length
*      54837     54936: gap of 100 bp
*      54937     55730: contig of 794 bp in length
*      55731     55830: gap of 100 bp
*      55831     56616: contig of 786 bp in length
*      56617     56716: gap of 100 bp
*      56717     57498: contig of 782 bp in length
*      57499     57598: gap of 100 bp
*      57599     58381: contig of 783 bp in length
*      58382     58481: gap of 100 bp
*      58482     59250: contig of 769 bp in length
*      59251     59350: gap of 100 bp
*      59351     60139: contig of 789 bp in length
*      60140     60239: gap of 100 bp
*      60240     61030: contig of 791 bp in length
*      61031     61130: gap of 100 bp
*      61131     61917: contig of 787 bp in length
*      61918     62017: gap of 100 bp
*      62018     62811: contig of 794 bp in length

Query Match      68.0% Score 17: DB 66: Length 69900;
Best Local Similarity 100.0% Pred. No. 17;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2      cgtcaggtcaggtcag 18
Db      19825      CGTCAGTCCAGTCAG 19809

RESULT 20
AC069132
LOCUS      AC069132      151340 bp      DNA
DEFINITION      Homo sapiens chromosome 5 clone RP11-556L8, *** SEQUENCING IN
PROGRESS ***; 62 unordered pieces.
ACCESSION      AC069132
VERSION      AC069132.2      GI:8469032
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1      (bases 1 to 151340)
AUTHORS      Waterston,R.H.
TITLE      The sequence of Homo sapiens clone

```

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
2 (bases 1 to 151340)
Waterston, R.H.
Direct Submission
Submitted (18-MAY-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jun 12, 2000 this sequence version replaced gi:7924001.

COMMENT

----- Genome Center -----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1502: contig of 1502 bp in length
* 1503 1602: gap of unknown length
* 1603 2741: contig of 1139 bp in length
* 2742 2841: gap of unknown length
* 2842 4036: contig of 1195 bp in length
* 4037 4136: gap of unknown length
* 4137 5620: contig of 1484 bp in length
* 5621 5720: gap of unknown length
* 5721 7080: contig of 1360 bp in length
* 7081 7180: gap of unknown length
* 7181 8839: contig of 1659 bp in length
* 8840 8939: gap of unknown length
* 8940 9987: contig of 1048 bp in length
* 9988 10087: gap of unknown length
* 10088 11156: contig of 1069 bp in length
* 11157 11256: gap of unknown length
* 11257 12404: contig of 1148 bp in length
* 12405 12504: gap of unknown length
* 12505 13861: contig of 1357 bp in length
* 13862 13961: gap of unknown length
* 13962 15241: contig of 1280 bp in length
* 15242 15341: gap of unknown length
* 15342 17009: contig of 1668 bp in length
* 17010 17109: gap of unknown length
* 17110 18897: contig of 1788 bp in length
* 18898 18997: gap of unknown length
* 18998 20389: contig of 1392 bp in length
* 20390 20489: gap of unknown length
* 20490 22042: contig of 1553 bp in length
* 22043 22142: gap of unknown length
* 22143 24057: contig of 1915 bp in length
* 24058 24157: gap of unknown length
* 24158 26189: contig of 2032 bp in length
* 26190 26289: gap of unknown length
* 26290 27959: contig of 1670 bp in length
* 27960 28059: gap of unknown length
* 28060 30067: contig of 2008 bp in length
* 30068 30167: gap of unknown length
* 30168 32456: contig of 2289 bp in length
* 32457 32556: gap of unknown length
* 32557 34597: contig of 2041 bp in length
* 34598 34697: gap of unknown length
* 34699 36611: contig of 1914 bp in length
* 36612 36711: gap of unknown length
* 36712 38677: contig of 1966 bp in length
* 38678 38777: gap of unknown length
* 38779 40974: contig of 2197 bp in length
* 40975 41074: gap of unknown length
* 41075 42666: contig of 1592 bp in length
* 42667 42766: gap of unknown length
* 42767 44292: contig of 1526 bp in length
* 44293 44392: gap of unknown length
* 44393 46032: contig of 1640 bp in length
* 46033 46132: gap of unknown length
* 46133 48815: contig of 2683 bp in length

48816 48916 48915: gap of unknown length
* 48916 51058: contig of 2143 bp in length
* 51059 51158: gap of unknown length
* 51159 53094: contig of 1936 bp in length
* 53095 53194: gap of unknown length
* 53195 54724: contig of 1530 bp in length
* 54725 54824: gap of unknown length
* 54825 57019: contig of 2195 bp in length
* 57020 57119: gap of unknown length
* 57120 59736: contig of 2617 bp in length
* 59737 59836: gap of unknown length
* 59837 61510: contig of 1674 bp in length
* 61511 61610: gap of unknown length
* 61611 63247: contig of 1637 bp in length
* 63248 63347: gap of unknown length
* 63348 65262: contig of 1915 bp in length
* 65263 65362: gap of unknown length
* 65363 67493: contig of 2131 bp in length
* 67494 67593: gap of unknown length
* 67594 70740: contig of 3147 bp in length
* 70741 70841: gap of unknown length
* 70841 72256: contig of 1416 bp in length
* 72257 72356: gap of unknown length
* 72357 73556: contig of 3100 bp in length
* 73557 75556: gap of unknown length
* 75557 77328: contig of 1772 bp in length
* 77329 77428: gap of unknown length
* 77429 78851: contig of 2423 bp in length
* 78852 79951: gap of unknown length
* 79952 82877: contig of 2926 bp in length
* 82878 82978: gap of unknown length
* 82979 84859: contig of 1882 bp in length
* 84860 84959: gap of unknown length
* 84960 87856: contig of 2897 bp in length
* 87857 87956: gap of unknown length
* 87957 91677: contig of 3721 bp in length
* 91678 91777: gap of unknown length
* 91778 94536: contig of 2759 bp in length
* 94537 94636: gap of unknown length
* 94637 97377: contig of 2741 bp in length
* 97378 97477: gap of unknown length
* 97478 100981: contig of 3504 bp in length
* 100982 101081: gap of unknown length
* 101082 103958: contig of 2877 bp in length
* 103959 104058: gap of unknown length
* 104059 106803: contig of 2845 bp in length
* 106804 107003: gap of unknown length
* 107004 109662: contig of 2659 bp in length
* 109663 109762: gap of unknown length
* 109763 113672: contig of 3910 bp in length
* 113673 113772: gap of unknown length
* 113773 118353: contig of 4581 bp in length
* 118354 118453: gap of unknown length
* 118454 121541: contig of 3088 bp in length
* 121542 121641: gap of unknown length
* 121642 124122: contig of 2481 bp in length
* 124123 124222: gap of unknown length
* 124223 128114: contig of 3892 bp in length
* 128115 128214: gap of unknown length
* 128215 131684: contig of 3470 bp in length
* 131685 131784: gap of unknown length
* 131785 135575: contig of 3791 bp in length
* 135576 135675: gap of unknown length
* 135676 140587: contig of 4912 bp in length
* 140588 140687: gap of unknown length
* 140688 145142: contig of 4455 bp in length
* 145143 145242: gap of unknown length
* 145243 151340: contig of 6098 bp in length.

FEATURES
source

1. 151340
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"

BASE COUNT 43058 a 29029 c 29780 g 43356 t 6117 others
 ORIGIN

Query Match 68.0%; Score 17; DB 74; Length 151340;
 Best local Similarity 100.0%; Pred. No. 14;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cgtcaggtcaggtcag 18
 |||||||
 Db 93019 CGTCAGTGCAGTCAG 93035

RESULT 21
 AC068100 155204 bp DNA HTG 16-MAR-2001
 LOCUS Homo sapiens chromosome X clone RP11-812N12 map X, WORKING DRAFT
 DEFINITION
 AC068100 4 GT:13357469
 AC068100
 VERSION
 KEYWORDS
 SOURCE HTG: HTGS_PHASE1; HTGS_DRAFT.
 ORGANISM
 human.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 1 (bases 1 to 155204)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Bonuslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Campiano, A., Castle, A., Choquel, Y., Colangelo, M., Collins, S., Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Howland, J., J. C., J. I., Johnson, R., Jones, C., Kann, L., Karats, A., Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehotzky, J., Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., McGuire, A., McKernan, K., McPeckers, R., Meldrum, J., Menus, L., Mihova, T., Miranda, C., Mianga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange, S., Theodores, J., Tittel, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission

TITLE
 JOURNAL
 COMMENT
 Submitted (28-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 16, 2001 this sequence version replaced g1:12383851.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: MIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L7131
 Center clone name: 812_N12

Summary Statistics
 Sequencing vector: M13: M77815; 6% of reads
 Sequencing vector: Plasmid: n/a; 94% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 146541 bases at least Q40

Consensus quality: 150817 bases at least Q30
 Consensus quality: 152368 bases at least Q20
 Insert size: 165000; agarose-fp
 Insert size: 153204; sum-of-contigs
 Quality coverage: 4.5 in Q20 bases; agarose-fp
 Quality coverage: 4.8 in Q20 bp.
 NOTE: This is a 'working draft' sequence. It currently consists of 21 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 518: contig of 518 bp in length
 * 519 618: gap of 100 bp
 * 619 1733: contig of 1115 bp in length
 * 1734 1833: gap of 100 bp
 * 1834 2717: contig of 884 bp in length
 * 2718 2817: gap of 100 bp
 * 2818 5343: contig of 2526 bp in length
 * 5344 5443: gap of 100 bp
 * 5444 8556: contig of 3113 bp in length
 * 8557 8656: gap of 100 bp
 * 8657 12409: contig of 3753 bp in length
 * 12410 12509: gap of 100 bp
 * 12510 16818: contig of 4309 bp in length
 * 16819 16918: gap of 100 bp
 * 16919 18984: contig of 2066 bp in length
 * 18985 19084: gap of 100 bp
 * 19085 41469: contig of 22385 bp in length
 * 41470 41569: gap of 100 bp
 * 41570 48545: contig of 6976 bp in length
 * 48546 48645: gap of 100 bp
 * 48646 55246: contig of 6601 bp in length
 * 55247 55346: gap of 100 bp
 * 55347 62502: contig of 7156 bp in length
 * 62503 62602: gap of 100 bp
 * 62603 70135: contig of 7533 bp in length
 * 70136 70235: gap of 100 bp
 * 70236 78809: contig of 8574 bp in length
 * 78810 78909: gap of 100 bp
 * 78910 88980: contig of 10071 bp in length
 * 88981 89080: gap of 100 bp
 * 89081 100598: contig of 11518 bp in length
 * 100599 100698: gap of 100 bp
 * 100699 115295: contig of 14597 bp in length
 * 115296 115395: gap of 100 bp
 * 115396 128681: contig of 13286 bp in length
 * 128682 128781: gap of 100 bp
 * 128782 140301: contig of 11520 bp in length
 * 140302 140401: gap of 100 bp
 * 140402 150306: contig of 9905 bp in length
 * 150307 150406: gap of 100 bp
 * 150407 155204: contig of 4798 bp in length.
 Location/Qualifiers
 1. 155204
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="X"
 /clone="RP11-812N12"
 /clone_id="RP11-812N12 Human Male BAC"
 1. 518
 /note="assembly-fragment"
 clone_end:Sp6
 vector_side:left
 619..1733
 /note="assembly-fragment"
 1834..2717
 /note="assembly-fragment"
 2818..5343
 /note="assembly-fragment"

misc_feature

misc_feature

misc_feature

misc_feature


```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0759J14
----- Summary Statistics -----
Sequencing vector: M13: 100%
Sequencing reagent: plasmid: 0%
Chemistry: dye-primer ET: 100% of reads
Chemistry: dye-terminator Big Dye: 0% of reads
Assembly program: Phrap: version 0.990319
Consensus quality: 157456 bases at least Q40
Consensus quality: 161145 bases at least Q30
Consensus quality: 163068 bases at least Q20
Insert size: 173000: agarose-fp
Insert size: 166760: sum-of-contigs
Quality coverage: 4.08 in Q20 bases; sum-of-contigs
Quality coverage: 4.27 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1711: contig of 1711 bp in length
*
1712 1811: gap of unknown length
*
1812 6135: contig of 4324 bp in length
*
6136 6235: gap of unknown length
*
6236 8687: contig of 2452 bp in length
*
8688 8787: gap of unknown length
*
8788 11647: contig of 2860 bp in length
*
11648 11747: gap of unknown length
*
11748 13718: contig of 1971 bp in length
*
13719 13818: gap of unknown length
*
13819 18016: contig of 4198 bp in length
*
18017 18116: gap of unknown length
*
18117 21570: contig of 3454 bp in length
*
21571 21570: gap of unknown length
*
21671 24723: contig of 3053 bp in length
*
24724 24823: gap of unknown length
*
24824 32650: contig of 7827 bp in length
*
32651 32750: gap of unknown length
*
32751 46491: contig of 13741 bp in length
*
46492 46591: gap of unknown length
*
46592 60127: contig of 13536 bp in length
*
60128 60227: gap of unknown length
*
60228 74481: contig of 14254 bp in length
*
74482 74581: gap of unknown length
*
74582 90886: contig of 16305 bp in length
*
90887 90886: gap of unknown length
*
90887 104984: contig of 13998 bp in length
*
104985 105084: gap of unknown length
*
105085 127379: contig of 22295 bp in length
*
127380 127478: gap of unknown length
*
127480 147029: contig of 19549 bp in length
*
147029 147128: gap of unknown length
*
147129 168360: contig of 21232 bp in length.
*
Location/Qualifiers
1. 168360
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-759J14"
1. 1711
/note="assembly_name:Contig6"
1812. 6135
/note="assembly_name:Contig7"
6236. 8687
/note="assembly_name:Contig8

```

```

misc_feature      clone_end:SP6
                  vector_side:right"
                  8788..11647
                  /note="assembly_name:Contig9"
misc_feature      11748..13718
                  /note="assembly_name:Contig10"
misc_feature      13819..18016
                  /note="assembly_name:Contig11"
misc_feature      18117..21570
                  /note="assembly_name:Contig12"
misc_feature      21671..24723
                  /note="assembly_name:Contig13"
misc_feature      24824..32650
                  /note="assembly_name:Contig14"
misc_feature      32751..46491
                  /note="assembly_name:Contig15"
misc_feature      46592..60127
                  /note="assembly_name:Contig16"
misc_feature      60228..74481
                  /note="assembly_name:Contig17"
misc_feature      74582..90886
                  /note="assembly_name:Contig18"
misc_feature      90987..104984
                  /note="assembly_name:Contig19"
misc_feature      105085..127379
                  /note="assembly_name:Contig20"
misc_feature      127480..147028
                  /note="assembly_name:Contig21"
misc_feature      147129..168360
                  /note="assembly_name:Contig22"
BASE COUNT      48985 a 34329 c 34643 g 48785 t      1618 others
ORIGIN

```

```

Query Match      68.0%; Score 17; DB 71; Length 168360;
Best Local Similarity 100.0%; Pred. NO. 14;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2      cgtcaggtgcaggtcag 18
          |||
Db 153451 CGTCAGTGCAGTGCAG 153435

```

```

RESULT 23
AC060833/c      176379 bp      DNA      HTG      17-AUG-2000
LOCUS      Homo sapiens chromosome 4 clone RP11-556F3, WORKING DRAFT SEQUENCE,
DEFINITION      4 unordered pieces.
ACCESSION      AC060833
VERSION      AC060833.3 GI:9838272
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 176379)
AUTHORS      Waterston,R.H.
JOURNAL      The sequence of Homo sapiens clone
TITLE      Unpublished
REFERENCE      2 (bases 1 to 176379)
AUTHORS      Waterston,R.H.
JOURNAL      Direct Submission
TITLE      Submitted (20-APR-2000) Genome Sequencing Center, Washington
JOURNAL      University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL      MO 63108, USA
COMMENT      On Aug 17, 2000 this sequence version replaced gi:8980022.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0556F03

```

```

----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-terminator Big Dye; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 173730 bases at least Q40
Consensus quality: 174579 bases at least Q30
Consensus quality: 175046 bases at least Q20
Insert size: 17400; agarose-fp
Insert size: 177415; sum-of-contigs
Quality coverage: 6.66 in Q20 bases; agarose-fp
Quality coverage: 6.55 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1419: contig of 1419 bp in length
* 1420 1519: gap of unknown length
* 1520 10016: contig of 8497 bp in length
* 10017 10116: gap of unknown length
* 10117 79219: contig of 69103 bp in length
* 79220 79319: gap of unknown length
* 79320 176379: contig of 97060 bp in length.
FEATURES
source
1..176379
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/clone="RP11-556F3"
1..1419
/note="assembly_name:Contig5"
1520..10016
/note="assembly_name:Contig6"
clone_end:SP6
vector_side:right"
10117..79219
/note="assembly_name:Contig7"
79320..176379
/note="assembly_name:Contig8"
clone_end:T7
vector_side:left"
BASE COUNT      49569 a 33984 c 35497 g 57025 t      304 others
ORIGIN

```

```

Query Match      68.0%; Score 17; DB 72; Length 176379;
Best Local Similarity 100.0%; Pred. NO. 14;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      2      cgtcaggtgcaggtcag 18
          |||
Db 81954 CGTCAGTGCAGTGCAG 81938

```

```

RESULT 24
AC022781      200729 bp      DNA      HTG      26-SEP-2000
LOCUS      Mus musculus chromosome 11 clone RP23-199H17 map 11, WORKING DRAFT
DEFINITION      SEQUENCE, 37 unordered pieces.
ACCESSION      AC022781
VERSION      AC022781.5 GI:10305259
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 200729)

```

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE Mus musculus chromosome 11, clone RP23-199H17
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 200729)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Beckely, R., Beda, F.,
 Boguslavsky, L., Bouknight, B., Brown, A., Burkett, G., Castle, A.,
 Choe, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P.,
 DeArnell, K., Dewar, K., Domino, M., Doyle, M., Fensholt, J.,
 Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
 Garaya, S., Grant, G., Hagos, B., Heath, A., Horton, L.,
 Howland, J. C., Johnson, R., Jones, C., Kan, L., Karatas, A., Klein, J.,
 Landers, T., Lechoczek, J., Levine, R., Liu, C., Liu, G., Locke, K.,
 MacDonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
 McPheters, R., Meldrim, J., Meneses, L., Morrow, J., Naylor, J.,
 Norman, C. H., O'Connor, T., O'Donnell, P., Olivar, T. M., Peterson, K.,
 Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
 Roy, A., Santos, R., Severy, P., Spencer, B., Strange-Thomann, N.,
 Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Titrill, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
 Zimmer, A. and Zody, M.

TITLE Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
 JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT On Sep 26, 2000 this sequence version replaced gi:154653.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www.seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

Project Information
 Center Project name: 199_H.17
 Center Clone name: 15913

Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 180464 bases at least Q40
 Consensus quality: 190837 bases at least Q30
 Consensus quality: 194945 bases at least Q20
 Insert size: 206000; agarose-fp
 Insert size: 197129; sum-of-contigs
 Quality coverage: 3.9 in Q20 bases; agarose-fp
 Quality coverage: 4.1 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
 consists of 37 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

7022 7121: contig of 7021 bp in length
 7122 8310: contig of 1189 bp in length
 8311 8410: gap of 100 bp
 8411 9411: contig of 1001 bp in length
 9412 9511: gap of 100 bp
 9512 11581: contig of 2070 bp in length
 11582 11681: gap of 100 bp
 11682 13342: contig of 1661 bp in length
 13343 13442: gap of 100 bp
 13443 16024: contig of 2582 bp in length
 16025 16124: gap of 100 bp
 16125 17440: contig of 1316 bp in length
 17441 17540: gap of 100 bp
 17541 18922: contig of 1382 bp in length
 18923 19022: gap of 100 bp
 19023 21353: contig of 2331 bp in length

FEATURES

source

21354 21453: gap of 100 bp
 21454 23674: contig of 2221 bp in length
 23675 23774: gap of 100 bp
 23775 25868: contig of 2094 bp in length
 25869 25968: gap of 100 bp
 25969 27966: contig of 1998 bp in length
 27967 28066: gap of 100 bp
 28067 31051: contig of 2985 bp in length
 31052 31151: gap of 100 bp
 31152 33236: contig of 2085 bp in length
 33237 33336: gap of 100 bp
 33337 36285: contig of 2949 bp in length
 36286 36385: gap of 100 bp
 36386 39476: contig of 3091 bp in length
 39477 39576: gap of 100 bp
 39577 42584: contig of 3008 bp in length
 42585 42684: gap of 100 bp
 42685 46164: contig of 3480 bp in length
 46165 46264: gap of 100 bp
 46265 49973: contig of 3709 bp in length
 49974 50073: gap of 100 bp
 50074 55389: contig of 5316 bp in length
 55390 55489: gap of 100 bp
 55490 59283: contig of 3794 bp in length
 59284 59383: gap of 100 bp
 59384 62877: contig of 3494 bp in length
 62878 62977: gap of 100 bp
 62978 66182: contig of 5205 bp in length
 66183 68282: gap of 100 bp
 68283 73862: contig of 5580 bp in length
 73863 73962: gap of 100 bp
 73963 83038: contig of 9076 bp in length
 83039 83138: gap of 100 bp
 83139 87922: contig of 4684 bp in length
 87923 87922: gap of 100 bp
 87923 92086: contig of 4164 bp in length
 92087 92186: gap of 100 bp
 92187 97007: contig of 4821 bp in length
 97008 97107: gap of 100 bp
 97108 123018: contig of 25911 bp in length
 123019 123118: gap of 100 bp
 123119 129307: contig of 6189 bp in length
 129308 129407: gap of 100 bp
 129408 137163: contig of 7756 bp in length
 137164 137263: gap of 100 bp
 137264 147510: contig of 10247 bp in length
 147511 147610: gap of 100 bp
 147611 155502: contig of 7892 bp in length
 155503 155602: gap of 100 bp
 155603 168939: contig of 13337 bp in length
 168940 169039: gap of 100 bp
 169040 181336: contig of 12297 bp in length
 181337 181436: gap of 100 bp
 181437 195395: contig of 13959 bp in length
 195396 195495: gap of 100 bp
 195496 200729: contig of 5234 bp in length.

Location/Qualifiers

1. 200729

/organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="11"
 /map="11"
 /clone="RP23-199H17"
 /clone.lib="RPC1-23 Female Mouse BAC"
 1..7021
 /note="assembly-fragment"
 clone_end:SP6
 vector_side:left
 7122..8310
 /note="assembly-fragment"
 8411..9411
 /note="assembly-fragment"
 9512..11581
 misc_feature
 misc_feature
 misc_feature
 misc_feature

misc_feature /note="assembly-fragment"
11682..13342
misc_feature /note="assembly-fragment"
13443..16024
misc_feature /note="assembly-fragment"
16125..17440
misc_feature /note="assembly-fragment"
17541..18922
misc_feature /note="assembly-fragment"
19023..21353
misc_feature /note="assembly-fragment"
21454..23674
misc_feature /note="assembly-fragment"
23775..25868
misc_feature /note="assembly-fragment"
25969..27966
misc_feature /note="assembly-fragment"
28067..31051
misc_feature /note="assembly-fragment"
31152..33236
misc_feature /note="assembly-fragment"
33337..36285
misc_feature /note="assembly-fragment"
36386..39476
misc_feature /note="assembly-fragment"
39577..42584
misc_feature /note="assembly-fragment"
42685..46164
misc_feature /note="assembly-fragment"
46265..49973
misc_feature /note="assembly-fragment"
50074..55389
misc_feature /note="assembly-fragment"
55490..59283
misc_feature /note="assembly-fragment"
59384..62877
misc_feature /note="assembly-fragment"
62978..68182
misc_feature /note="assembly-fragment"
68283..73862
misc_feature /note="assembly-fragment"
73963..83038
misc_feature /note="assembly-fragment"

Query Match 68.0%; Score 17; DB 67; Length 200729;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cgtcaggtgcaggtcag 18
|||||
Db 69699 CGTCAGGTGCAGGTGACG 69715

RESULT 25
LOCUS CNS078T3 1044 bp DNA
DEFINITION T7 end of clone BB0A023D11 of library BB0AA from strain CBS 4732
ACCESSION AL434413
VERSION AL434413.1 GI:12217827
KEYWORDS STS.
SOURCE Pichia angusta.
ORGANISM Pichia angusta
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE
AUTHORS Blandin, G., Llorente, B., Malpertuy, A., Wincker, P., Artiguenave, F.,
and Dujon, B.
TITLE Genomic Exploration of the Hemiascomycetous Yeasts: 13. Pichia
angusta
JOURNAL FEBS Lett. 487 (1), 76-81 (2000)
PUBMED 11152888
REFERENCE 2 (bases 1 to 1044)

AUTHORS Souciet, J.-L., Aigle, M., Artiguenave, F., Blandin, G.,
Boloitin-Pukuhara, M., Bon, E., Brothier, P., Casaregola, S.,
de-Montigny, J., Dujon, B., Durrens, P., Lepoint, A., Llorente, B.,
Malpertuy, A., Neugeilse, C., Ozier-Kalogeropoulos, O., Potier, S.,
Saurin, W., Tekala, F., Toffano-Nioche, C., Weslowski-Louvel, M.,
Wincker, P., and Weissbach, U.
TITLE Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
yeast species for molecular evolution studies(1)
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
PUBMED 11152876
REFERENCE 3 (bases 1 to 1044)
AUTHORS
TITLE Genoscope.
JOURNAL Direct Submission
COMMENT Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
secref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This STS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxi,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
Location/Qualifiers
1..1044
/organism="Pichia angusta"
/strain="CBS 4732"
/db_xref="taxon:4905"
/clone="BB0A023D11"
/clone_1lb="BB0AA"
/note="end : 17"
complement(<1..>999)
/note="similar to Saccharomyces cerevisiae ORF YGL026c [
TRP5 ; tryptophan synthase 1]"
/evidence="not-experimental"

BASE COUNT 245 a 311 c 228 g 257 t 3 others

Query Match 64.0%; Score 16; DB 53; Length 1044;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ggtcaggtgcaggtcag 22
|||||
Db 546 GGTGACGTGACGACG 531

RESULT 26
LOCUS AF237379 8326 bp DNA
DEFINITION Xenopus laevis neurofilament protein (NF-M(2)) gene, complete cds.
ACCESSION AF237379
VERSION AF237379.1 GI:11139431
KEYWORDS STS.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE
AUTHORS Roosta, J.R., Gervasi, C., and Szaro, B.G.
TITLE Structure, biological activity of the upstream regulatory sequence,
and conserved domains of a middle molecular mass neurofilament gene
of Xenopus laevis
JOURNAL Brain Res. Mol. Brain Res. 82 (1-2), 35-51 (2000)
PUBMED 11042356
REFERENCE 2 (bases 1 to 8326)
AUTHORS Roosta, J.R., and Szaro, B.G.
TITLE Direct Submission

JOURNAL Submitted (18-FEB-2000) Biological Sciences, University at Albany,
State University of New York, 1400 Washington Avenue, Albany, NY
12222, USA

FEATURES

Location/Qualifiers

1..8326

/organism="Xenopus laevis"

/db_xref="taxon:8345"

TATA_signal

1498..1503

5'UTR

1524..1563

exon

1524..2598

/number=1

join(1524..2598,3533..3657,6507..8326)

mRNA

/product="neurofilament protein"

gene

1564..8326

/gene="NF-M(2)"

join(1564..2598,3533..3657,6507..8088)

/gene="NF-M(2)"

/codon_start=1

/product="neurofilament protein"

/protein_id="AAG31697.1"

/db_xref="GI:11139432"

/translation="MSYLETIGNPSYRMATPTSYSTSTPSSGFRSOSMSRSP
TTISYKRTNLAPRYNSNDLSPSSAFNGDLKOVSEKEDLOGINDFVYIE
KHPLEQONKLEAFIALROKSGYSGLEYEDMRRLRYNLELIGDKOYIDAF
EHLEDDLOKLEKEBEARITRPTENATIAKKMDTSLVYELDGHSQNMLOTEEV
IKNNHEEYGLLAQIASITVERKDFOTLASKRLTYSOLEGHSQNMLOTEEV
FNRRYAKLFEAEQNKARSAKDIAEYRQLHSKRYELVETITKIVREKNDIE
EHHGDLTSYQETIOOLDNELRGTKEMSRHLREVODLLNVMALDIEIAARKLLEG
EETRFSGVGTGSAIHRPTISTRVQARAEPKRVQHKVEELIETKVEDD
RSMEMALAAADPEFALGFEGEEGEEGEEGEEGEEGEEGEEGEEGEEGEEGEE
EAEKFE
EERKEGEGEDAEVEETKAEBEDEBEDEBEDEBEDEBEDEBEDEBEDEBEDEBEDE
EGEGEGEDKDEBEKEGEAEAEAGGEEETTEAQAEVVEETITKIVREKAEV
OETAEDEETQAEBOKDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE
EKGEIDEGTKEGAEDE
TORDDCKEDTAVNGVEGEVESEDKAVVNGDDSPSEDEKESKGGQIVYITKKI
ETTTSRGEDEGTITKTSVITQKVEHEETVQEKLVSTKKVEKVTSHAVKVEETDK"
2599..3532

intron

/gene="NF-M(2)"

/number=1

3533..3657

/gene="NF-M(2)"

intron

3659..6507

/gene="NF-M(2)"

/number=2

6507..8326

/gene="NF-M(2)"

/number=3

8089..8326

/gene="NF-M(2)"

3'UTR

/gene="NF-M(2)"

BASE COUNT 2631 a 1560 c 2161 g 1974 t

ORIGIN

Query Match 64.0%; Score 16; DB 8; Length 8326;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 caagtcagcagtcacga 20
|||||

DB 3060 CAGTCGCGTCACGA 3075

RESULT 27
AE004050 10536 bp DNA BCT 06-JUL-2000
LOCUS Xylella fastidiosa, section 196 of 229 of the complete genome.
DEFINITION AE004050 AE003849
ACCESSION AE004050.1 GI:9107594
VERSION
KEYWORDS
SOURCE Xylella fastidiosa.
ORGANISM Xylella fastidiosa

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

Bacteria: Proteobacteria; gamma subdivision; Xanthomonas group; Xylella.

1 (bases 1 to 10536)

Silvestri,M.L., Siqueira,M.J., de Souza,A.A., de Souza,A.P., Terezi,M.F., Truffi,D., Tsai,S.M., Tsubako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S., Vettore,A.L., Zago,M.A., Zatz,M., Melandri,J. and Setubal,J.C

The genome sequence of the plant pathogen Xylella fastidiosa. The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing and Analysis, Sao Paulo, Brazil

Nature 406 (6792), 151-157 (2000)

2 (bases 1 to 10536)

Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Agencio,M., Alvaerenga,R., Alves,L.M.C., Araya,J.E., Bala,G.S., Baptista,C.S., Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Britton,M.R.S., Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M., Carrier,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R., Costa-Neto,C.M., Coutinho,L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorri,H., Facincani,A.P., Ferreira,A.J.S., Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C., Fritche,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,H.S., Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F., Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A., Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N., Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L., Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Mitacca,E.C., Miyaki,C.Y., Montello-Vitello,C.B., Moon,D.H., Nagel,F.G., Nascimento,A.L.T.O., Netto,L.E.S., Nhani,J.R., Nobrega,F.G., Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C., Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira,Jr.,H.A., Pequeiro,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V., de M. Rosa,A.J., de Rosa,Jr.,F.E., de Sa,R.R., da Silva,A.M., Silva Sawsaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva Jr.,W.A., da Silveira,J.F., Silvestri,M.L., Siqueira,M.J., de Souza,A.A., de Souza,A.P., Terezi,M.F., Truffi,D., Tsai,S.M., Tsubako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S., Vettore,A.L., Zago,M.A., Zatz,M., Melandri,J. and Setubal,J.C.

Direct Submission

Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP 13083-970, Brazil

REMARK

FEATURES

source

Location/Qualifiers

1..10536

/organism="Xylella fastidiosa"

/db_xref="taxon:2371"

1296..2339

/gene="XF2409"

1296..2339

/gene="XF2409"

/note="similar to SPIP39369 (percent identity: 47 %/query alignment coverage: 96.8 %/subject alignment coverage: 99.4 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder/Start codon shift: 2523"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AA85208.1"

/db_xref="GI:9107595"

/translation="MPALGVLAVDGLCKSSGGRYNYEATTAAMLAHEHESLQAO
YGEPLIRIVGILPETOARAISSQACRVGIEVGEHKGDIYGVTHALOGAERPVVIF
SAVYSGHADGFTIDORASMLNVAASRAKTEFLVFGDMVFTAAKSRPGLAHYLFK
DPSNALCROPILRKDLQVSTVAEVLDAEHDAFLDLANKVOREIHIVSPWINKDR
TDDICAFKMOAEVROVQYVYTPDDINTDQKDKTKRIAEVLDAARLRGIEVHF
VDRVSKMVTGDEVFVGVSGFWFSAANSAMAKHETSLVYRGRLADERQRLNSLR
QRITDDPQGVGA"

311..2484

/gene="XF2410"

2311..2484

/gene="XF2410"

/note="hypothetical protein; identified by sequence

```

similarity: putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA085209.1"
/db_xref="GI:9107596"
/translation="MILSKSARDAVLQGRUDVGRWGMHAAEGNDMSPVALNA
VALKCCCLPLG"
complement(3181..5514)
/gene="XF2411"
complement(3181..5514)
/gene="XF2411"
/note="hypothetical protein; identified by sequence
similarity: putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA085210.1"
/db_xref="GI:9107597"
/translation="MLKKLKVNHVGLLRNGVPTPTPPGVPVTLFYAENGKCTTVA
NLRVTPHGKDEVLQKARTIDPDDEPCTELLMDGNTDOKYTLHHGHTGTPDL
IEDAFVDDNYSQGEVHPHARQKFLQFVGAESVKLEREIVLAKETIADENRITKE
EQRIRASKNPPIADFTLIPVDAEAKIALRTQREAVNAPALLQRAHAPDLPHE
CNDAFPTISQSTPHDIRADAEITVQAHFQHQDPGIEAMVQGFRENAQGGPCG
QEVATALIKAYETVFNFAKFAEMKNATLGSVATPEACELEKIDARLOCDNACVA
AMKDQDLSPPFNPRATATFTLRTVALALASQKHRLPDCVGTODKADWRLLE
ELRAOMDAVNAAYEINRITNFRKTLAENMPALDASITATLETGIVROSTEVQAT
VYQAKAKYVQLQKREVALGRTSDSPGHNVLSBQDKRTLALAFLLAYAPDIL
VGSVVLDDPWCSTDMTRNRTMESIALVNOGVVYLSDHAFILRLDLLAGPY
NKVSVNHHIKRTKNDSQIVSDADLDTICOSPMLRYAQQVTFVSGTIGTQEVAS
ALRPLVEGFLKHFAPPLRLQDLSLQMSIDIRAKTDDSPVLAKPYVDITLKNFA
VQSHHDSKSLSPMNDGRLROYAFMALLIYGLSPH"
complement(5615..5800)
/gene="XF2412"
complement(5615..5800)
/gene="XF2412"
/note="hypothetical protein; identified by sequence
similarity: putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA085211.1"
/db_xref="GI:9107598"
/translation="MIFGATQTNARRRDSGTRHAACTGWAARSRLPPDQVRNHLIP
KPHVYISSIPITGTM"
6084..6512
/gene="XF2413"
6084..6512
/gene="XF2413"
/note="hypothetical protein; identified by sequence
similarity: putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA085212.1"
/db_xref="GI:9107599"
/translation="WVGDDVPTADTRAAAGACEPSIKNIATSRGVHLLCAGMRLM
PGCCCLGNGPMKGSVYKSSVGLVLRGSCQOYEAARGPAPRTFGKYMSALML
YRMWCSRHCLSTRTEGMGATSGDAGPRGRGMLKGRG"
complement(6509..8314)
/gene="XF2414"
complement(6509..8314)
/gene="XF2414"
/note="hypothetical protein; identified by sequence
similarity: putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA085213.1"
/db_xref="GI:9107600"
/translation="MPPKRHOVADADTKENLTVTTEHCIFGCVIFPLDGGSGOTFG
FSFELAGLMLVLAWTIADVRRFRIQVAPLPLKNSFYVVAIGILLTLTLMRAEO

```

tRNA

```

/note="found by tRNAscan"
/product="tRNA-Thr"
complement(8365..8440)
/gene="tRNA-Thr-2"
complement(8457..8648)
/gene="XF2415"
complement(8457..8648)
/gene="XF2415"

```

CDS

```

/note="hypothetical protein; identified by sequence
similarity: putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA085214.1"
/db_xref="GI:9107601"
/translation="MEVPRVDPQPPILKSHLCATVALLQSLPSPDPAIILMRAP
EAHTKQSSALSRSPSTAM"
complement(8921..9871)
/gene="XF2416"
complement(8921..9871)
/gene="XF2416"

```

gene

CDS

```

/note="similar to GI:5915671 (percent identity: 64 %/query
alignment coverage: 99.4 %/subject alignment coverage:
99.4 %); identified by sequence similarity: putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="drug tolerance protein"
/protein_id="AA085215.1"
/db_xref="GI:9107602"
/translation="MKVLLANRRCAGVDAIEIVKRTIDMLGPIYVREYVHNR
VVDLKGKAIIVFELHVPDQATVIFSAHGVSAVROAQRGLKYDADTCPLVTV
HLDVABHCRGTGDMILIGHAGPEVEGGMQDGERGRYILENIDVDTLLVAPD
HILATQPTTISYDPTNIDALRRPPTTGGPKKNDICVATOKRQDAVRLAECIL
VLVGSPPNSNSNRSLSELAQRGVAASYTIDSAEIDPAWVDKHHIGVTAGASAPVYL
VDGVLARLIEGATISVSEHSGPESWVFLPRLRLQILDV"

```

```

Query Match      64.0%; Score 16; DB 1; Length 10536;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      2 cytcaggtcaggtca 17
DB      1995 CGTCAGGTGACAGGTCA 2010

```

RESULT 28

AE005241

LOCUS

```

DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 65

```

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

```

1 Escherichia coli O157:H7 EDL933.
2 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
3 Escherichia.
4 (bases 1 to 12242)
5 Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D.,
6 Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J.J., Kirkpatrick, H.A.,
7 Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,

```

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 Grobbeck, E.J., Davis, N.W., Lim, A., Dimantanta, E., Potamou, K.,
 Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
 Welch, R.A. and Blattner, F.R.
 Genome Sequence of enterohaemorrhagic *Escherichia coli* O157:H7
 Nature 409 (819), 529-533 (2001)
 21074935
 11206551
 2 (bases 1 to 12242)
 Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D.,
 Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
 Posal, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
 Grobbeck, E.J., Davis, N.W., Lim, A., Dimantanta, E., Potamou, K.,
 Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
 Welch, R.A. and Blattner, F.R.
 Direct Submission
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 Submitted (22-OCT-2000) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 Location/Qualifiers
 1. 12242
 /organism="Escherichia coli O157:H7 EDL933"
 /strain="EDL933"
 /serotype="O157:H7"
 /db_xref="taxon:155864"
 /note="enterohaemorrhagic"
 complement(146. .952)
 /gene="rna"
 /note="20755"
 complement(146. .952)
 /gene="rna"
 /function="enzyme: Macromolecule degradation: Degradation
 of RNA"
 /note="Residues 1 to 268 of 268 are 99.25 pct identical to
 residues 1 to 268 of 268 from *Escherichia coli* K-12 Strain
 MG1655: B0611"
 /codon_start=1
 /transl_table=11
 /product="RNase I, cleaves phosphodiester bond between any
 two nucleotides"
 /protein_id="AAG54946.1"
 /db_xref="GI:12513510"
 /translation="MKAFMRNALIVSLPSSANAVALQAKQYEDFRVYLAISWQ
 TGRGSOHNRNREDECRQETTTKADFLVHGLMPKIPSVARGVDERMMWEG
 GATRPVLPNLEPARASRMCSPETGSLSETRAKLSEVMPGAGRSCLEREYAKHGACF
 GPDPAVFGTIVWNLQEIKESEKFLADNKGTVSRDPAFAFAKSMCKENYKAVKL
 TCGNPAYLLEIOWSITKADAINAPLSANSFLPQHPHNGCKTFVLDKAGY"
 complement(1066. .2529)
 /gene="ytds"
 /note="20756"
 complement(1066. .2529)
 /gene="ytds"
 /function="putative membrane: Not classified"
 /note="Residues 1 to 487 of 487 are 100.00 pct identical
 to residues 1 to 487 of 487 from *Escherichia coli* K-12
 strain MG1655: B0612"
 /codon_start=1
 /transl_table=11
 /product="putative a membrane protein"
 /protein_id="AAG54947.1"
 /db_xref="GI:12513511"
 /translation="MSIAKDIWKLAIPLVYVGVFLPVPQMPPOAMHFAVAVAM
 IYGMLEPAPATISFIANTYTCVGSNTLTPAKELADPAFAKQALQKMGAGCSST
 TWLVGAFITPAGLEVSGLRRIALFLVFKPGKRTLLGVAIVLIDLLAFPPSNT
 ARTGTGTPVPIKINLPLEKSPNDPSARKIGYLMWMMVISTLSMSEVGAANVL
 GLEPVSKLAGIOWLQWFLCPLPVGVILLITAPWISVLYLPEITHSSEVATAGDE
 LKTMGLATREWTLLGLVLLSLGLWFGSEVINATAVGILAVLSIMLAIHVPWMDTR
 YNSAMNTLMTLVVMANGLTRSGFIMFAGCTSTHEGSPNATVIVLVFPAH
 YLFASATATATMLPVIIAVGKXIPGVEMEDLCTLLVLSIGIMGLTPYATGPGVITY
 GCGYKSKDYRMLGALFEGVITYSMLLVGMPLAMWN"
 complement(2580. .3458)
 /gene="cItg"
 /note="20757"
 complement(2580. .3458)
 /gene="cItg"

gene
 CDS
 /function="orf: Unknown function"
 /note="Residues 1 to 292 of 292 are 98.97 pct identical to
 residues 1 to 292 of 292 from *Escherichia coli* K-12 Strain
 MG1655: B0613"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAG54948.1"
 /db_xref="GI:12513512"
 /translation="MSMPATIKTKTALSLIDEVALLGMRAMTEVNSPKGLVDR
 INGAHKDMLDEDFHRSALAIQGMIPRTFERPACSEMAPEVILGRTIGACGDM
 FRATPAVNTNKGSTFSLGILCATIGLNDLPTTVCTSPASFCRGLDRELRTN
 NSDITAGORIVQDLGTGARGAEAGYPLVINHAIPHYLLIDOGIDELALDLILL
 LMAINDGTVAASNGEGGLRWLQREAOQTLLQGGIRTPADLDYLLQFQPRECIERNLSP
 GGSADLLITWFLAQI"
 complement(3433. .3984)
 /gene="ytdu"
 /note="20758"
 complement(3433. .3984)
 /gene="ytdu"
 /function="orf: Unknown function"
 /note="Residues 1 to 183 of 183 are 99.45 pct identical to
 residues 1 to 183 of 183 from *Escherichia coli* K-12 Strain
 MG1655: B0614"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAG54949.1"
 /db_xref="GI:12513513"
 /translation="MHILPELASHHAVSIPELVSRDROKARQHWLKRHPVLYSTF
 VYAPRKDESEYTRIFNHCYTRALALAKQWQLOEOPALVASGPGMUSIAAPAR
 DKLATTELEHNPUGRLMDIDLTPTEGELISRDYSLPPRCLCEQSAAVCAVARGKT
 HOLVDLNLMEHPLNDVDAQVNV"
 complement(3988. .5520)
 /gene="cItf"
 /note="20759"
 complement(3988. .5520)
 /gene="cItf"
 /function="enzyme: Central intermediary metabolism: Pool,
 multipurpose conversions of intermedi. met'm"
 /note="Residues 1 to 510 of 510 are 99.41 pct identical to
 residues 1 to 510 of 510 from *Escherichia coli* K-12 Strain
 MG1655: B0615"
 /codon_start=1
 /transl_table=11
 /product="citrate lyase alpha chain"
 /protein_id="AAG54950.1"
 /db_xref="GI:12513514"
 /translation="MTOKIRDSORQERYAANRRACEDLAFQNSPKQTYOAEARR
 KLCANLEAIRRSIGIDGMITYSEHARNGDLYNMMVDVIAKMGFKNLIASSISD
 CHAPLVEHTRGVVTRITTSGLRPLAEISRGILAEVQIHSRGVHLVQSGELSD
 DVAFVGVSCDEFGNANGYTGKACCGSLGYMVDADNAKQVWLEILLPYHPNPASI
 EODQVDLIVKDVODAKIGAGATRMNTNRELLIRASADAVLVNNGYEGESMOT
 GTGASLAVTLFLEDDKMSRDIRADFLAGTTAVMDLHEKGLIRKLLDQSPSHAA
 OSLARNPHIEISANQYANMGSKASGADVVDVVLASLEIDPQNNVLYGSDVYLG
 ASGHCDDPAISALSIIYAPVPRGRIPLVNDVNTCTTPGSSVDLLVDGCIANPAR
 PELARRLOEAGIKVYSIEWLERARLLTGEPOPLETDRVAVAVRYRDSYDVVHCV
 KE"
 complement(5531. .6454)
 /gene="cItg"
 /note="20760"
 complement(5531. .6454)
 /gene="cItg"
 /function="enzyme: Central intermediary metabolism: Pool,
 multipurpose conversions of intermedi. met'm"
 /note="Residues 1 to 307 of 307 are 99.34 pct identical to
 residues 1 to 307 of 307 from *Escherichia coli* K-12 Strain
 MG1655: B0616"
 /codon_start=1
 /transl_table=11
 /product="citrate lyase beta chain (acyl lyase subunit)"
 /protein_id="AAG54951.1"
 /db_xref="GI:12513515"


```

Equivalent to M.tuberculosis Rv3584 (MTV024.02), 63.4% identity in 175 aa overlap. Contains N-terminal signal sequence with appropriately positioned PS00013 Prokaryotic membrane lipoprotein lipid attachment site"
/transl_table=11
/product="putative lipoprotein"
/protein_id="CAA22686.1"
/db_xref="GI:4154038"
/db_xref="SPTREMBL:Q9ZBM7"
/translation="MSRRKISLPALTRAVVLGFTIMASVLCGCAGQISQTATDEP AVNNRNVLLNMLLRDIRIQAACGTDLGSGRVDLMVLAINNSPVYDRLVSITSDI GTVALNGCTGTGTMGLFETSGEQRILKKPPLOSNINIAKAIVTLAPITNGLTNYETF NEFKAGGNVAVPVPSAGLAPRGRT"
714..746
/gene="MlCB1450.02"
/note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"
1428..2463
/gene="rada"
1428..2463
/gene="rada"
/note="MlCB1450.03, probable rada pseudogene, len: 1033 bp, Equivalent to M.tuberculosis Rv3585, (MTV024.03), probable DNA repair protein RADA homolog"
/codon_start=1
/pseudo
/transl_table=11
/product="putative rada pseudogene"
/db_xref="PID:e1370359"
/complement(172)..5344
/note="identical to EMBL:MLANTM_M14341 M.leprae 65 kd antigen, from 12 to 3613"
/complement(2559..3208)
/note="RLEP1"
2832..3170
/gene="MlCB1450.04"
2832..3170
/gene="MlCB1450.04"
/note="MlCB1450.04", CDS in RLEP element, len: 112 aa; identical to CDS in other RLEP elements e.g. TR:O32969 (MLCP22.36C)"
/codon_start=1
/transl_table=11
/product="hypothetical protein MlCB1450.04"
/protein_id="CAA22688.1"
/db_xref="GI:4154039"
/translation="MTTPRPQHDMHTKPLPRGANNYPHTHACIDLAIFSAQVSPDW HHQHDQAPRTDMLTCAALIYSTAAKHTEHKRKQAVSHPTTKTPHQHSKTQDQSLP PTTTHIGCS"
complement(3504..5129)
/gene="groEL-2"
complement(3504..5129)
/gene="groEL-2"
/note="MlCB1450.05c, groEL-2, 60 kD chaperonin, len: 541 aa; identical to Ch62_MyCLE, and highly similar to many e.g. Ch60_BACSU 60 kD chaperonin (543 aa), Jasta scores; opt: 2217 z-score: 2520.3 E(): 0, 63.3% identity in 539 aa overlap. Equivalent to M.tuberculosis Ch62_MYCU RV00440 (94.8% identity in 541 aa overlap). Contains PS00296 chaperonins cpn60 signature and Pfam match to entry PF00118 cpn60_TCPI, TCP-1 /cpn60 chaperonin family"
/codon_start=1
/transl_table=11
/product="60 kD chaperonin"
/protein_id="CAA22689.1"
/db_xref="GI:4154040"
/translation="MAKTIAYDEEARGLERGINSLADAVKVTLGPKRNVLLEKKW APITNDGSIKTELIEDPYEKIGAEIKVEAKKTDVAGSTTAIVLDAOLVKKEG LRVVAAGANPLCIKRGEIKAVKYKVTILLDKAEKVETEKOATAAISAGDOSIGDL AEMDNAGNGVIYVESNRFGILOELITBGMFEKGYSIGSYFYTAEEQNAVLEERY LAIVSSVSRMLDPLELVKVIQAGSKSLTIADVEGAESLTVNWK IGTGRSVANKV PGGDGRMRKLDMALITIGAQVISEVVGITLNTDNLSTLGKARKVVMKDEFITVEGG

```

misc_feature	complement(3564..5066)		/gene="groEL-2"		/note="Pfam match to entry PF00118 cpn60_TCP1, TCP-1	
misc_feature	complement(3388..3923)		/cpn60 chaperonin family, score 778.80, E-value 2.2e-230"			
gene	complement(5338..5907)		/gene="MLCB1450..06c"		complement(5338..5907)	
CDS	/gene="MLCB1450..06c"		/note="MLCB1450..06c, probable transcriptional regulator, len: 189 aa; similar to hypothetical proteins from M. tuberculosis and S. coelicolor, and equivalent to M.tuberculosis putative transcriptional regulator RV0067c (68.8% identity in 189 aa overlap). Contains probable helix-turn-helix motif at aa 35-56 (Score 1794, +5.30 SD) and Pfam match to entry PF00440 tetr, Bacterial regulatory proteins, tetr family"		/codon_start=1	
	/transl_table=1		/product="putative transcriptional regulator"		/protein_id="CA422690.1"	
	/db_xref="GI:415041"		/db_xref="SPTRMBL:092BM6"		/translation="MAQPARPLRDAAARNRARIIGVAFATFEGISVPIDELAFRAA	
	/translation="MAQPARPLRDAAARNRARIIGVAFATFEGISVPIDELAFRAA		VAGVYRHRFPKKEALCAAVIGDMPLHVDGYYLLKSAPGALFTYIRSLYLHMG		TRGVLDALAGAGGIDILIGVAPDAEDAFVLTSLDLYAOTAGTARDIVGEVNSII	
	VCCQAMEYVNSALERVYDVVDIGLRAAR"					
Query Match	64.0%: Score 16; DB 3; Length 38065;					
Best Local Similarity	100.0%: Pred. No. 65;					
Matches	16; Conservative		0; Mismatches		0; Indels	
	0; Gaps		0			
QY	5 caagtcgacgctcagca 20					
14193	CAGGTGCAGCTCAGCA 14208					
RESULT	30					
AC091120	39520 bp DNA		HTG		30-MAR-2001	
LOCUS	Leishmania major chromosome 2 clone L6541 strain F1ed11n, ***					
DEFINITION	SEQUENCING IN PROGRESS ***, 3 ordered pieces.					
ACCESSION	AC091120.1		GI:13489160			
VERSION	AC091120.1		GI:13489160			
KEYWORDS	HTG; HTGS_PHASE2.					
SOURCE	Leishmania major					
ORGANISM	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.					
REFERENCE	1 (bases 1 to 39520)					
AUTHORS	Myler,P.J., Sisk,E., Ruiz,J., Cosenza,P., Cruz,A., Stuart,K., Robertson,L., Vogt,C., Marty,A., Cawthra,J., Seyler,A., Munden,H., Nelson,S., Mack,J. and Pentony,M.					
TITLE	Direct Submission					
JOURNAL	Submitted (30-MAR-2001) Seattle Biomedical Research Institution, 4					
COMMENT	Nickerson Street, Seattle, WA 98109-1651, USA					
	* NOTE: This is a 'working draft' sequence. It currently					
	* consists of 3 contigs. Gaps between the contigs					
	* are represented as runs of N. The order of the pieces					
	* is believed to be correct as given, however the sizes					
	* of the gaps between them are based on estimates that have					
	* provided by the submittor.					
	* This sequence will be replaced					
	* by the finished sequence as soon as it is available and					
	* the accession number will be preserved.					
	1 27709: contig of 27709 bp in length					
	* gap of unknown length					

misc_feature	complement(3564..5066) /gene="groEL-2" /note="Pfam match to entry PF00118 cpn60_TCP1, TCP-1/cpn60 chaperonin family, score 778.80, E-value 2.2e-230"	
misc_feature	complement(3888..3923) /gene="groEL-2"	
gene	/note="PS00296 Chaperonins cpn60 signature" complement(538..5907) /gene="MLCB1450..06c" complement(5338..5907) /gene="MLCB1450..06c" /note="MLCB1450..06c, probable transcriptional regulator, len: 189 aa; similar to hypothetical proteins from M. tuberculosis and S. coelicolor, and equivalent to M.tuberculosis putative transcriptional regulator RV0067c (68.8% identity in 189 aa overlap). Contains probable helix-turn-helix motif at aa 35-56 (Score 1794, +5.30 SD) and Pfam match to entry PF00440 tetr, Bacterial regulatory proteins, tetr family"	
CDS	/codon_start=1 /transl_table=1 /product="putative transcriptional regulator" /protein_id="CA422690.1" /db_xref="GI:415041" /db_xref="SPTREMBL:092BM6" /translation="MAQPARPLRDAAARNRARIIGVAFEFATEGISVPIDEIARARA VAGVIRHFPKKEALCAAVIGDMPLHVDGYYLLKSAPGALFTYIRSLYLHMG TRGVLDALAGAGGIDILGVAPDAEDAFVLTSLDLYAOTAGTARDVGVEVSSII VCCQAMEYVNSALERVTVDDVVDGLRAAR"	
Query Match	64.0%: Score 16; DB 3; Length 38065; Best Local Similarity 100.0%; Pred. No. 65;	
Matches	16; Conservative	0; Mismatches 0; Indels 0; Gaps 0
QY	5	caagtcgcaagtcagca 20
DB	14193	CAGGTGCAGCTCAGCA 14208
RESULT	30	
AC091120	39520 bp	DNA HTG \ 30-MAR-2001
LOCUS	Leishmania major chromosome 2 clone L6541 strain F1ed1ln, ***	
DEFINITION	SEQUENCING IN PROGRESS ***, 3 ordered pieces.	
ACCESSION	AC091120.1	GI:13489160
VERSION	HTG; HTGS_PHASE2.	
KEYWORDS	Leishmania major.	
SOURCE	Leishmania major	
ORGANISM	Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.	
REFERENCE	1 (bases 1 to 39520) Myler,P.J., Sisk,E., Ruiz,J., Cosenza,P., Cruz,A., Stuart,K., Robertson,L., Vogt,C., Marty,A., Cawthra,J., Seyler,A., Munden,H., Nelson,S., Mack,J. and Pentony,M. Direct Submission Submitted (30-MAR-2001) Seattle Biomedical Research Institution, 4 Nickerson Street, Seattle, WA 98109-1651, USA	
TITLE	* NOTE: This is a 'working draft' sequence. It currently	
JOURNAL	* consists of 3 contigs. Gaps between the contigs	
COMMENT	* are represented as runs of N. The order of the pieces	
	* is believed to be correct as given, however the sizes	
	* of the gaps between them are based on estimates that have	
	* provided by the submittor.	
	* This sequence will be replaced	
	* by the finished sequence as soon as it is available and	
	* the accession number will be preserved.	
	1	27709: contig of 27709 bp in length
	* gap of unknown length	

```

*      27710      38388: contig of 10679 bp in length
*      38389      39520: gap of unknown length
*      Location/Qualifiers
*          1. 39520
*              /organism="Leishmania major"
*              /strain="Friedlin"
*              /db_xref="taxon:5664"
*              /chromosome="2"
*              /clone="L6541"
*              /c1one="L6541"
BASE COUNT      6965 a 12678 c 12387 g 7487 t      3 others
ORIGIN

Query Match
Best Local Similarity 64.0%; Score 16; DB 78; Length 39520;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5      caagtcaggtcagca 20
Db      1027      CAGGTCAGGTCAGCA 1042

RESULT 31
AC004508      71230 bp      DNA      PRI      30-MAR-1998
LOCUS      Homo sapiens chromosome 21, p1 clone LBL#3 (LBLN H09), complete
sequence.
ACCESSION      AC004508      L35684      L35685      L35686      L35687      AC000939      AC000938      AC000940
AC000937      L81573      L35689      L35690      L35691      L35692      L35693      L35694      L35695
L35696      L35697      L35698      L35699      L42077      L35700      L35701
AC004508.1      GI:2996630
VERSION      HTG.
KEYWORDS      human.
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 71230)
AUTHORS      Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,
Davis, C.A., Kadner, K., Miguel, T., Pitluck, S., Pollard, M.,
Rojeski, H., Subramanian, S. and Martin, C.H.
TITLE      Sequencing of human chromosome 21
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 71230)
AUTHORS      Rieke, D.O.
TITLE      Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
JOURNAL      Unpublished
REFERENCE      3 (bases 1 to 71230)
AUTHORS      Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,
Davis, C.A., Kadner, K., Miguel, T., Pitluck, S., Pollard, M.,
Rojeski, H., Subramanian, S. and Martin, C.H.
TITLE      Direct Submission
JOURNAL      Submitted (30-MAR-1998) Human Genome Center, DOE Joint Genome
Institute, Lawrence Berkeley National Laboratory, MS 74-157,
Berkeley, CA 94720, U.S.A.
COMMENT      Sequence submitted by:
DOE Joint Genome Institute.
FEATURES
source
1. 71230
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/clone="LBL#3"
/clone="LBLN H09"
/clone="LBLN H09"
/complement(472..788)
/rpt_family="MIR1a"
1408..1743
/rpt_family="Alu"
4057..4515
/feature="99% identity zhp2d01.r1"
/db_xref="dbEST:W86547"
complement(4506..4622)
misc_feature

misc_feature
/feature="GRAIL 2 excellent exon, frame 2"
/feature="100% identity zhp2d01.s1"
/db_xref="dbEST:W86451"
complement(4646..5114)
/standard_name="SHC-52452"
/feature="100% identity"
/db_xref="dbSTS:G33332"
complement(5148..5285)
/rpt_family="Alu"
complement(5788..5902)
/rpt_family="MER44C"
complement(5974..6182)
/rpt_family="MER44C"
complement(6153..6473)
/rpt_family="Alu"
complement(6485..6565)
/rpt_family="MER44C"
complement(6779..6989)
/rpt_family="MER46"
complement(7235..8312)
/rpt_family="L1"
complement(8410..8569)
/rpt_family="L1"
8577..8987
/rpt_family="THER1"
complement(9020..9511)
/rpt_family="L1"
9872..10296
/rpt_family="THER1"
10297..11909
/rpt_family="MSTAR"
11929..12309
/rpt_family="THER1"
12868..13131
/rpt_family="Alu"
complement(13267..13635)
/rpt_family="L1"
join(14992..15122,21091..21151,22200..22263,25113..25251,
33644..33721)
/standard_name="YJ99d04"
14993..15116
/feature="98% identity R74138"
/feature="GRAIL 2 excellent exon, frame 0"
join(15038..15122,21091..21151,22209..22263,25113..25251,
33644..33885)
/standard_name="ZU55D03"
/feature="98% identity AA402132"
complement(15991..16164)
/rpt_family="MER42"
complement(16400..16743)
/rpt_family="MIR2B2"
complement(16761..17073)
/rpt_family="Alu"
16768..16787
/feature="97% identity"
/rpt_type="tandem"
19581..19626
/feature="GRAIL 2 excellent exon, frame 0"
20435..20765
/rpt_family="Alu"
23936..24290
/rpt_family="Alu"
24273..24292
/feature="A120"
/rpt_type="tandem"
25115..25247
/feature="GRAIL 2 excellent exon, frame 2"
26872..26965
/rpt_family="MIR"
29019..29360
repeat_region

```

```

repeat_region /rpt_family="Alu"
                29356..31388
repeat_region /rpt_family="L1"
                31440..32072
repeat_region /rpt_family="L1"
                complement(32120..32230)
repeat_region /rpt_family="L1"
                32581..32624
                /note="(AC)22"
                /rpt_type=tandem
                /rpt_unit=AC
repeat_region complement(32668..32776)
                /rpt_family="MER3"
                32932..33120
                /note="86% identity 2h52f10.r1"
                /db_xref="dbEST:w78940"
                complement(33133..33199)
                /rpt_family="MER3"
repeat_region complement(33644..33883)
                /note="99% identity nt66401.s1"
                /db_xref="dbEST:AA515600"
                33648..33737
                /note="GRAIL 2 excellent exon, frame 2"
                35260..35660
                /rpt_family="MLT1"
                36246..36533
                /rpt_family="Alu"
                complement(39299..39589)
                /rpt_family="Alu"
                complement(39781..39931)
                /rpt_family="Alu"
                complement(41134..41588)
                /rpt_family="MLT2"
                41621..41706
                /rpt_family="MLT1"
                41748..41771
                /note="(GTTT)6"
                /rpt_type=tandem
                /rpt_unit=GTTT
                complement(41757..42397)
                /rpt_family="L1"
                42935..43106
                /rpt_family="Alu"
                complement(44760..44802)
                /note="GRAIL 2 excellent exon, frame 2"
                complement(49449..49589)
                /rpt_family="MIR2"
                complement(51869..52337)
                /rpt_family="MLT1"
                52773..53082
                /rpt_family="Alu"
                <53503..>67215
                /gene="MX2"
                /note="interferon-regulated resistance GTP-binding protein
                MXB (p78-related protein)"
                /evidence=experimental
                join(<53503..53668,54302..54494,56530..56664,58923..59077,
                67077..>67215)
                /gene="MX2"
                /note="interferon-regulated resistance GTP-binding protein
                MXB (p78-related protein)"
                /codon_start=2
                /evidence=experimental
                /product="MX2"
                /protein_id="AAC08448.1"
                /db_xref="GI:2996631"
                /translation="FOQOPPGFVPPQMFPPNMOGAEKDAFLAKDFNLTLLNPP
                PGNRSOPRAMGPNENLYSOYEQKVRPCIDILSLRALGVEDLALPAIVIGDOSSG
                KSSVLEALSGVALPGSGIVRCPLVYKLRKCEAMAGRIYRNLELDDGQVEK
                EIRKAQVWAGNGRGISHELISLETSPVPDLTIIDPGITRVAVDNQPRDGLGDIK
                ALIKYIQRQDTIMLVVPCNDIATTEALSMHEVDPEDRIT"
repeat_region 57922..58215
                /rpt_family="Alu"

```

```

repeat_region 59599..59877
                /rpt_family="Alu"
repeat_region 60428..60810
                /rpt_family="THE1"
repeat_region 61246..61366
                /rpt_family="Alu"

Query Match          64.0%; Score 16; DB 85; Length 71230;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 gtcagtgacgtcag 18
    ||| ||| ||| ||| |||
Db 40455 GTCAGGTGACGTGAG 40470

```

```

RESULT 32
AC002088/c LOCUS AC002088
DEFINITION Human BAC clone CTB-13p7 from 7p15-p21, complete sequence.
ACCESSION AC002088
VERSION AC002088.1 GI:2085779
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 85633)
AUTHORS Waterston,B. and Scheet,P.
JOURNAL The sequence of H. sapiens BAC clone CTB-13p7
REFERENCE 2 (bases 1 to 85633)
AUTHORS Unpublished
JOURNAL
TITLE Submitted (13-MAY-1997)
AUTHORS Waterston,B.
JOURNAL Direct Submission
TITLE Submitted (03-FEB-2000) Department of Genetics, Washington
AUTHORS University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
JOURNAL 4 (bases 1 to 85633)
REFERENCE 4 (bases 1 to 85633)
AUTHORS Waterston,B.
JOURNAL Direct Submission
TITLE Submitted (04-FEB-2000) Department of Genetics, Washington
AUTHORS University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
JOURNAL ----- Genome Center
COMMENT Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_RG013p07

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION: The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7> or send an E-mail to egreen@nhgri.nih.gov

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).
 VECTOR: pBeloBAC11
 Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of CTB-13P7; actual end is at 85633 of CTB-13P7. The orientation of this clone is unknown.

FEATURES
 This clone contains SWS2365 (NID:q1254783).

source	Location/Qualifiers
repeat_region	1..85633 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="7" /clone="CTB-13P7" /map="7p15-p21" complement(3420..3446)
repeat_region	/rpt_family="L1" complement(4016..4306)
repeat_region	/rpt_family="ALU" complement(6902..6942)
repeat_region	/rpt_family="L1" 13108..13401 complement(16600..16982)
repeat_region	/rpt_family="ALU" complement(17300..17375)
repeat_region	/rpt_family="ALU" complement(17475..17502)
repeat_region	/rpt_family="L1" 17503..17584
repeat_region	/rpt_family="ALU" 17625..17838
repeat_region	/rpt_family="ALU" 17979..18078
repeat_region	/rpt_family="L1" complement(18407..18454)
repeat_region	/rpt_family="L1" complement(24978..25240)
repeat_region	/rpt_family="ALU" 25293..25377
repeat_region	/rpt_family="L1" complement(25503..25778)
repeat_region	/rpt_family="L1" complement(25779..26070)
repeat_region	/rpt_family="ALU" complement(26076..26884)
repeat_region	/rpt_family="L1" complement(26887..27177)
repeat_region	/rpt_family="ALU" complement(27242..27327)
repeat_region	/rpt_family="L1" complement(27436..27612)
repeat_region	/rpt_family="L1" complement(28177..28229)
repeat_region	/rpt_family="L1" complement(28230..28517)
repeat_region	/rpt_family="ALU" complement(28518..28570)
repeat_region	/rpt_family="L1" 31781..32083
repeat_region	/rpt_family="MER" 32102..32130
repeat_region	/rpt_family="MER" complement(34983..35002)
repeat_region	/rpt_family="L1" complement(35680..35721)
repeat_region	/rpt_family="L1" complement(38336..38360)
repeat_region	/rpt_family="L1" complement(39172..39464)
repeat_region	/rpt_family="ALU" complement(39570..39698)
repeat_region	/rpt_family="ALU" complement(39954..40230)
repeat_region	/rpt_family="L1" 40007..40195
repeat_region	/rpt_family="L1" complement(40341..40901)
repeat_region	/rpt_family="L1" 40345..40893
repeat_region	/rpt_family="L1" complement(41771..47939)
repeat_region	/rpt_family="L1" complement(47950..48026)
repeat_region	/rpt_family="MER" complement(48181..48428)
repeat_region	/rpt_family="MER" 48233..48439
repeat_region	/rpt_family="MER" complement(48619..48655)
repeat_region	/rpt_family="MER" complement(48694..48890)
repeat_region	/rpt_family="MER" 48819..48974
repeat_region	/rpt_family="MER" 49030..49292
repeat_region	/rpt_family="MER" 49440..49757
repeat_region	/rpt_family="MER" complement(49570..49873)
repeat_region	/rpt_family="MER" 51098..51155
repeat_region	/rpt_family="ALU" 51181..51230
repeat_region	/rpt_family="ALU" complement(51563..51830)
repeat_region	/rpt_family="ALU" complement(52400..52441)
repeat_region	/rpt_family="L1" complement(53064..53085)
repeat_region	/rpt_family="L1" complement(53087..53384)
repeat_region	/rpt_family="ALU" complement(53410..53435)
repeat_region	/rpt_family="L1" complement(53436..53545)
repeat_region	/rpt_family="ALU" complement(53554..53709)
repeat_region	/rpt_family="ALU" complement(60972..60992)
repeat_region	/rpt_family="L1" complement(64802..64824)
repeat_region	/rpt_family="L1" complement(66753..67258)
repeat_region	/rpt_family="MER" complement(72167..72185)
repeat_region	/rpt_family="L1" 72879..73153
repeat_region	/rpt_family="ALU" complement(82252..82468)
repeat_region	/rpt_family="MER" 82281..82473
repeat_region	/rpt_family="MER" 82600..82926
repeat_region	/rpt_family="MER" complement(83480..83780)


```

----- Summary Statistics -----
Sequencing vector: M13, 93%
Sequencing vector: plasmid: 7%
Chemistry: Dye-primer ET; 93% of reads
Chemistry: Dye-terminator Big Dye; 7% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 134469 bases at least Q40
Consensus quality: 137723 bases at least Q30
Consensus quality: 139595 bases at least Q20
Insert size: 162000; agarose-fp
Quality coverage: 3.63 in Q20 bases; agarose-fp
Quality coverage: 3.95 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1571: contig of 1571 bp in length
1572 1671: gap of unknown length
1672 2958: contig of 1287 bp in length
2959 3058: gap of unknown length
3059 4454: contig of 1396 bp in length
4455 4554: gap of unknown length
4555 5616: contig of 1062 bp in length
5617 7860: contig of 2144 bp in length
7861 7960: gap of unknown length
7961 11316: contig of 3356 bp in length
11317 14833: contig of 3417 bp in length
14834 14934: gap of unknown length
14934 18349: contig of 3415 bp in length
18349 22186: gap of unknown length
22186 25357: gap of unknown length
25357 25457: contig of 3071 bp in length
25458 28970: contig of 3513 bp in length
28971 31930: contig of 2860 bp in length
31931 32030: gap of unknown length
32031 35417: contig of 3387 bp in length
35418 35518: gap of unknown length
35518 39647: contig of 4130 bp in length
39648 43858: contig of 4111 bp in length
43859 43959: gap of unknown length
43960 49718: contig of 5660 bp in length
49719 54322: gap of unknown length
54323 54422: contig of 4604 bp in length
54423 61625: gap of unknown length
61626 61725: contig of 7203 bp in length
61726 65884: contig of 4159 bp in length
65885 71895: gap of unknown length
71896 71995: contig of 5911 bp in length
71996 78239: gap of unknown length
78240 83879: contig of 6244 bp in length
83880 83979: gap of unknown length
83980 90865: contig of 6886 bp in length
90866 90965: gap of unknown length
90966 98738: contig of 7773 bp in length
98739 98838: gap of unknown length
98839 110905: contig of 12067 bp in length
110906 111005: gap of unknown length
111006 124936: contig of 13931 bp in length

```

```

FEATURES             * 124937 125036: gap of unknown length
                        * 125037 147129: contig of 22093 bp in length.
SOURCE
1. 147129
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="2"
  /clone="RP11-92F20"
  1. 1571
    /note="assembly_name:Contig14"
    1672..2958
    /note="assembly_name:Contig15"
    3059..4454
    /note="assembly_name:Contig16"
    4455..5616
    /note="assembly_name:Contig17"
    5717..7860
    /note="assembly_name:Contig19"
    7961..11316
    /note="assembly_name:Contig20"
    11417..14833
    /note="assembly_name:Contig21"
    14934..18348
    /note="assembly_name:Contig22"
    18449..22186
    /note="assembly_name:Contig23"
    22287..25357
    /note="assembly_name:Contig24"
    25458..28970
    /note="assembly_name:Contig25"
    29071..31930
    /note="assembly_name:Contig26"
    32031..35417
    /note="assembly_name:Contig27"
    35518..39647
    /note="assembly_name:Contig28"
    39748..43858
    /note="assembly_name:Contig29"
    43959..49618
    /note="assembly_name:Contig30"
    49719..54322
    /note="assembly_name:Contig31"
    54423..61625
    /note="assembly_name:Contig32"
    61726..65884
    /note="assembly_name:Contig33"
    65985..71895
    /note="assembly_name:Contig34"
    71996..78239
    /note="assembly_name:Contig35"
    78340..83879
    /note="assembly_name:Contig36
    clone_end:17
    vector_side:right"
    misc_feature
    /note="assembly_name:Contig37"
    83980..90865
    /note="assembly_name:Contig38"
    90966..98738
    /note="assembly_name:Contig39"
    98839..110905
    /note="assembly_name:Contig39"
    111006..124936
    /note="assembly_name:Contig40"
    125037..147129
    /note="assembly_name:Contig41"
    2633 others
BASE COUNT  44181 a 28681 c 26793 g 44841 t
ORIGIN

```

```

Query Match          64.0%; Score 16; DB 74; Length 147129;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 caggtcaggtcagca 20

```

Db 6359 CAGGTGAGGTGAGCA 6354

RESULT 36
AC027810/C
LOCUS AC027810 155628 bp DNA HTG 26-MAY-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-357J12 map 18, WORKING DRAFT
SEQUENCE 14 unordered pieces.
AC027810
AC027810.3 GI:8082020
VERSION
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 155628)
2 (bases 1 to 155628)
Unpublished

REFERENCE
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Bogdanavich, L., Boukhgalter, B., Brown, A., Burkett, G.,
Campiano, A., Castle, A., Choquel, Y., Colangelo, M., Collins, S.,
Corry, A., Cooke, P., Deaile, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferrel, P., Fitzhugh, M., Gage, D.,
Galaan, J., Gardina, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Holland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karas, A.,
Klein, J., Larocque, K., Lamazares, R., Landers, T., Lebecky, J.,
Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPherson, R.,
Meldrum, J., Menus, L., Mihova, T., Miranda, C., Miengo, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, R., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talmas, J.,
Tessier, S., Theodore, J., Tittel, A., Travers, M., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT
Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced g1:7684514.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L9097
Center clone name: 357-J-12
Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 146315 bases at least Q40
Consensus quality: 151241 bases at least Q30
Consensus quality: 153082 bases at least Q20
Insert size: 163000; agarose-1p
Quality coverage: 4.2 in Q20 bases; agarose-1p
Quality coverage: 4.4 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1068: contig of 1068 bp in length
* 1069 1168: gap of 100 bp
* 1169 2717: contig of 1549 bp in length
* 2718 2817: gap of 100 bp
* 2818 6040: contig of 3223 bp in length
* 6041 6140: gap of 100 bp
* 6141 10201: contig of 4061 bp in length
* 10202 10301: gap of 100 bp
* 10302 14615: contig of 4314 bp in length
* 14616 14715: gap of 100 bp
* 14716 21086: contig of 6371 bp in length
* 21087 21186: gap of 100 bp
* 21187 25767: contig of 4581 bp in length
* 25768 25867: gap of 100 bp
* 25868 32500: contig of 6633 bp in length
* 32501 32600: gap of 100 bp
* 32601 38651: contig of 6051 bp in length
* 38652 38751: gap of 100 bp
* 38752 50922: contig of 12171 bp in length
* 50923 51022: gap of 100 bp
* 51023 71014: contig of 19992 bp in length
* 71015 71114: gap of 100 bp
* 71115 93563: contig of 2249 bp in length
* 93564 93663: gap of 100 bp
* 93664 123080: contig of 29617 bp in length
* 123081 123180: gap of 100 bp
* 123181 155628: contig of 32448 bp in length.
Location/Qualifiers
1. 155628
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone.lib="RP11-357J12"
/clone.lib="RP11-357J12" Human Male BAC"
1. 1068
/note="assembly-fragment"
1169. 2717
/note="assembly-fragment"
2818. 6040
/note="assembly-fragment"
6141. 10201
/note="assembly-fragment"
10302. 14615
/note="assembly-fragment"
14716. 21086
/note="assembly-fragment"
21187. 25767
/note="assembly-fragment"
25868. 32500
/note="assembly-fragment"
32601. 38651
/note="assembly-fragment"
38752. 50922
/note="assembly-fragment"
51023. 71014
/note="assembly-fragment"
71115. 93563
/note="assembly-fragment"
93664. 123080
/note="assembly-fragment"
123181. 155628
/note="assembly-fragment"
clone_end:SP6
vector_side:left"

BASE COUNT 37665 a 38986 c 40608 g 37067 t 1302 others
ORIGIN

TITLE
JOURNAL
COMMENT

Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, D., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N.,
McCarthy, M., McEwan, P., McGuire, A., McKernan, K., McPheters, R.,
Meltrin, J., Meunier, L., Mihova, T., Miranda, C., Mlepea, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Ollivar, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisanil, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testafave, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zalnoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 3, 2001 this sequence version replaced gi:7705184.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L8890
Center clone name: 2006.N.8

----- Summary Statistics

Sequencing vector: M13; M77815; 47% of reads
Sequencing vector: Plasmid; n/a; 53% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 159666 bases at least Q40
Consensus quality: 159912 bases at least Q30
Consensus quality: 160439 bases at least Q20
Insert size: 157000; agarose-1p
Insert size: 161043; sum-of-coverage
Quality coverage: 10.5 in Q20 bases; agarose-1p
Quality coverage: 10.2 in Q20.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 840: contig of 840 bp in length
* 841 940: gap of 100 bp
* 941 1652: contig of 712 bp in length
* 1653 1752: gap of 100 bp
* 1753 2845: contig of 1093 bp in length
* 2846 2945: gap of 100 bp
* 2946 4039: contig of 1094 bp in length
* 4040 4139: gap of 100 bp
* 4140 7025: contig of 2886 bp in length
* 7026 7125: gap of 100 bp
* 7126 10842: contig of 3717 bp in length
* 10843 10942: gap of 100 bp
* 10943 18373: contig of 7431 bp in length
* 18374 18473: gap of 100 bp
* 18474 28208: contig of 9735 bp in length
* 28209 28308: gap of 100 bp
* 28309 48596: contig of 20288 bp in length
* 48597 48696: gap of 100 bp
* 48697 76612: contig of 27916 bp in length
* 76613 76712: gap of 100 bp
* 76713 162043: contig of 85331 bp in length.
Location/Qualifiers
1. 162043

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="CTD-2006N8"
/clone_id="CTD Human BAC"
1. 840
/note="assembly-fragment"
misc_feature
941. 1652
/note="assembly-fragment"
misc_feature
1753. 2845
/note="assembly-fragment"
misc_feature
2946. 4039
/note="assembly-fragment"
misc_feature
4140. 7025
/note="assembly-fragment"
misc_feature
7126. 10842
/note="assembly-fragment"
misc_feature
10943. 18373
/note="assembly-fragment"
misc_feature
18474. 28208
/note="assembly-fragment"
misc_feature
28309. 48596
/note="assembly-fragment"
misc_feature
48697. 76612
/note="assembly-fragment"
misc_feature
76713. 162043
/note="assembly-fragment"
BASE COUNT 44171 a 35675 c 34546 g 46650 t 1001 others
ORIGIN

Query Match 64.0%; Score 16; DB 71; Length 162043;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 caggtcaggtcagca 20
|||||
Db 82976 CAGGTCCAGTCCAGCA 82991

RESULT 39
AC068538/C
LOCUS
DEFINITION
29 unordered pieces.
ACCESSION
AC068538
VERSION
AC068538.5 GI:9795677
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 175459)
Walterson, R.H.
TITLE
The sequence of Homo sapiens clone
JOURNAL
Unpublished
2 (bases 1 to 175459)
Walterson, R.H.
REFERENCE
Direct Submission
Submitted (03-MAY-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 11, 2000 this sequence version replaced gi:9690368.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Project Information
Center project name: H_NH0141G04

Sequencing vector: M13; 100%

Sequencing vector: plasmid; 0%
 Chemistry: Dye-primer ET; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 158616 bases at least Q40
 Consensus quality: 162423 bases at least Q30
 Consensus quality: 164318 bases at least Q20
 Insert size: 179000; agarose-fp
 Insert size: 172659; sum-of-ctrls
 Quality coverage: 4.51 in Q20 bases; agarose-fp
 Quality coverage: 4.87 in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently consists of 29 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1355: contig of 1355 bp in length
 1356 1455: gap of unknown length
 1456 3082: contig of 1627 bp in length
 3083 3183: gap of unknown length
 3183 4263: contig of 1081 bp in length
 4263 4364: gap of unknown length
 4364 5627: contig of 1263 bp in length
 5627 5727: gap of unknown length
 5727 7534: contig of 1807 bp in length
 7534 9202: contig of 1569 bp in length
 9202 9303: gap of unknown length
 9303 11061: contig of 1759 bp in length
 11061 14017: gap of unknown length
 14017 14118: contig of 2856 bp in length
 14118 16928: contig of 2811 bp in length
 16928 17029: gap of unknown length
 17029 19922: contig of 2894 bp in length
 19922 20023: gap of unknown length
 20023 23986: contig of 3964 bp in length
 23986 24087: gap of unknown length
 24087 27864: contig of 3778 bp in length
 27864 27965: gap of unknown length
 27965 31696: contig of 3731 bp in length
 31696 31796: gap of unknown length
 31796 34671: contig of 2876 bp in length
 34671 34772: gap of unknown length
 34772 39460: contig of 4688 bp in length
 39460 39560: gap of unknown length
 39560 43565: contig of 4006 bp in length
 43565 43666: gap of unknown length
 43666 47319: contig of 3553 bp in length
 47319 47418: gap of unknown length
 47418 53986: contig of 6568 bp in length
 53986 54087: gap of unknown length
 54087 61187: contig of 7100 bp in length
 61187 61287: gap of unknown length
 61287 66332: contig of 5045 bp in length
 66332 66431: gap of unknown length
 66431 72447: contig of 6016 bp in length
 72447 72548: gap of unknown length
 72548 80701: contig of 8154 bp in length
 80701 89408: gap of unknown length
 89408 89509: contig of 8607 bp in length
 89509 100141: gap of unknown length
 100141 100240: contig of 10632 bp in length
 100240 109860: contig of 9620 bp in length
 109860 120505: gap of unknown length
 120505 120605: contig of 10545 bp in length
 120605 135065: gap of unknown length
 135065 135066: contig of 14460 bp in length

FEATURES	
source	
misc_feature	135066 135165: gap of unknown length
misc_feature	135166 153777: contig of 18612 bp in length
misc_feature	153778 153877: gap of unknown length
misc_feature	153878 175459: contig of 21582 bp in length.
misc_feature	location/Qualifiers
misc_feature	1. 175459
misc_feature	/organism="Homo sapiens"
misc_feature	/db_xref="taxon:9606"
misc_feature	/chromosome="2"
misc_feature	/clone="RP11-141G4"
misc_feature	1. 1355
misc_feature	/note="assembly_name:Contig9"
misc_feature	1456. 3082
misc_feature	/note="assembly_name:Contig11"
misc_feature	3183. 4263
misc_feature	/note="assembly_name:Contig13"
misc_feature	4364. 5626
misc_feature	/note="assembly_name:Contig14"
misc_feature	5727. 7533
misc_feature	/note="assembly_name:Contig15"
misc_feature	7634. 9202
misc_feature	/note="assembly_name:Contig16"
misc_feature	clone_end:17
misc_feature	vector_side:left"
misc_feature	9303. 11061
misc_feature	/note="assembly_name:Contig17"
misc_feature	11162. 14017
misc_feature	/note="assembly_name:Contig18"
misc_feature	14118. 16928
misc_feature	/note="assembly_name:Contig19"
misc_feature	17029. 19922
misc_feature	/note="assembly_name:Contig20"
misc_feature	20023. 23986
misc_feature	/note="assembly_name:Contig21"
misc_feature	24087. 27864
misc_feature	/note="assembly_name:Contig22"
misc_feature	27965. 31695
misc_feature	/note="assembly_name:Contig23"
misc_feature	31796. 34671
misc_feature	/note="assembly_name:Contig24"
misc_feature	34772. 39459
misc_feature	/note="assembly_name:Contig25"
misc_feature	39560. 43565
misc_feature	/note="assembly_name:Contig26"
misc_feature	43666. 47318
misc_feature	/note="assembly_name:Contig27"
misc_feature	47419. 53986
misc_feature	/note="assembly_name:Contig28"
misc_feature	54087. 61186
misc_feature	/note="assembly_name:Contig29"
misc_feature	61287. 66331
misc_feature	/note="assembly_name:Contig30"
misc_feature	66432. 72447
misc_feature	/note="assembly_name:Contig31"
misc_feature	clone_end:sp6
misc_feature	vector_side:right"
misc_feature	72548. 80701
misc_feature	/note="assembly_name:Contig32"
misc_feature	80802. 89408
misc_feature	/note="assembly_name:Contig33"
misc_feature	89509. 100140
misc_feature	/note="assembly_name:Contig34"
misc_feature	100241. 109860
misc_feature	/note="assembly_name:Contig35"
misc_feature	109961. 120505
misc_feature	/note="assembly_name:Contig36"
misc_feature	120606. 135065
misc_feature	/note="assembly_name:Contig37"
misc_feature	135166. 153777
misc_feature	/note="assembly_name:Contig38"
misc_feature	153878. 175459
misc_feature	/note="assembly_name:Contig39"
BASE COUNT	47828 a 38505 c 38363 g 47961 t 2802 others

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 1026: contig of 1026 bp in length
* 1027 1126: gap of 100 bp
* 1127 2310: contig of 1184 bp in length
* 2311 2410: gap of 100 bp
* 2411 3533: contig of 1123 bp in length
* 3534 3633: gap of 100 bp
* 3634 4974: contig of 1341 bp in length
* 4975 5074: gap of 100 bp
* 5075 6187: contig of 1113 bp in length
* 6188 6287: gap of 100 bp
* 6288 7383: contig of 1096 bp in length
* 7384 7483: gap of 100 bp
* 7484 8508: contig of 1025 bp in length
* 8509 8608: gap of 100 bp
* 8609 9759: contig of 1191 bp in length
* 9800 9899: gap of 100 bp
* 9900 11275: contig of 1376 bp in length
* 11276 11375: gap of 100 bp
* 11376 12804: contig of 1429 bp in length
* 12805 12904: gap of 100 bp
* 12905 14246: contig of 1342 bp in length
* 14247 14346: gap of 100 bp
* 14347 15544: contig of 1198 bp in length
* 15545 15644: gap of 100 bp
* 15645 17181: contig of 1537 bp in length
* 17182 17281: gap of 100 bp
* 17282 19392: contig of 2111 bp in length
* 19393 19492: gap of 100 bp
* 19493 21710: contig of 2218 bp in length
* 21711 21810: gap of 100 bp
* 21811 24149: contig of 2339 bp in length
* 24150 24249: gap of 100 bp
* 24250 26685: contig of 2436 bp in length
* 26686 26785: gap of 100 bp
* 26786 28327: contig of 1542 bp in length
* 28328 28427: gap of 100 bp
* 28428 31535: contig of 3108 bp in length
* 31536 31635: gap of 100 bp
* 31636 34165: contig of 2510 bp in length
* 34166 34265: gap of 100 bp
* 34266 37131: contig of 2866 bp in length
* 37132 37231: gap of 100 bp
* 37232 40119: contig of 2888 bp in length
* 40120 40219: gap of 100 bp
* 40220 43234: contig of 3015 bp in length
* 43235 43334: gap of 100 bp
* 43335 46125: contig of 2791 bp in length
* 46126 46225: gap of 100 bp
* 46226 48748: contig of 2533 bp in length
* 48749 48848: gap of 100 bp
* 48849 53250: contig of 4402 bp in length
* 53251 53350: gap of 100 bp
* 53351 57172: contig of 3822 bp in length
* 57173 57272: gap of 100 bp
* 57273 60281: contig of 3009 bp in length
* 60282 60381: gap of 100 bp
* 60382 63484: contig of 3103 bp in length
* 63485 63584: gap of 100 bp
* 63585 67025: contig of 3441 bp in length
* 67026 67125: gap of 100 bp
* 67126 70475: contig of 3350 bp in length
* 70476 70575: gap of 100 bp
* 70576 74638: contig of 4063 bp in length
* 74639 74738: gap of 100 bp
* 74739 78257: contig of 3519 bp in length
* 78258 78357: gap of 100 bp
* 78358 83359: contig of 5002 bp in length
* 83360 83459: gap of 100 bp
* 83460 86986: contig of 3527 bp in length

```

FEATURES

Source

```

* 86987 87086: gap of 100 bp
* 87087 90187: contig of 3101 bp in length
* 90188 90287: gap of 100 bp
* 90288 94735: contig of 4448 bp in length
* 94736 94835: gap of 100 bp
* 94836 99768: contig of 4933 bp in length
* 99769 99868: gap of 100 bp
* 99869 104278: contig of 4410 bp in length
* 104279 104378: gap of 100 bp
* 104379 110808: contig of 6430 bp in length
* 110809 110908: gap of 100 bp
* 110909 116478: contig of 5570 bp in length
* 116479 116578: gap of 100 bp
* 116579 122443: contig of 5865 bp in length
* 122444 122543: gap of 100 bp
* 122544 126331: contig of 3788 bp in length
* 126332 126431: gap of 100 bp
* 126432 131374: contig of 4943 bp in length
* 131375 131474: gap of 100 bp
* 131475 138177: contig of 6703 bp in length
* 138178 138277: gap of 100 bp
* 138278 144292: contig of 6015 bp in length
* 144293 144392: gap of 100 bp
* 144393 150325: contig of 5933 bp in length
* 150326 150425: gap of 100 bp
* 150426 156168: contig of 5743 bp in length
* 156169 156268: gap of 100 bp
* 156269 163498: contig of 7230 bp in length
* 163499 163598: gap of 100 bp
* 163599 173145: contig of 9547 bp in length
* 173146 173245: gap of 100 bp
* 173246 185492: contig of 12247 bp in length.

```

FEATURES

Source

```

1..185492
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-676D22"
/clone_11p="RPCT-11 Human Male BAC"
1..1026
/note="assembly_fragment"
1127..2310
/note="assembly_fragment"
2411..3533
/note="assembly_fragment"
3634..4974
/note="assembly_fragment"
5075..6187
/note="assembly_fragment"
6288..7383
/note="assembly_fragment"
7484..8508
/note="assembly_fragment"
8609..9759
/note="assembly_fragment"
9900..11275
/note="assembly_fragment"
11376..12804
/note="assembly_fragment"
12905..14246

```

Query Match

Best Local Similarity 100.0%; Pred. No. 46;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 caggtgcagtcagca 20
 |||

DB 43512 CAGGTGCAGTCAGCA 43497

RESULT 42

AC064835

LOCUS AC064835 202828 bp DNA HTG 07-JUL-2000
 DEFINITION Homo sapiens chromosome 16 clone RP11-673p17, WORKING DRAFT
 ACCESSION AC064835
 VERSION AC064835.3 GI:7839924
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 202828)
 AUTHORS Waterston, R.H.
 TITLE The sequence of Homo sapiens clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 202828)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (22-APR-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 COMMENT On May 16, 2000 this sequence version replaced g1:7712280.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 ----- Project Information -----
 Center project name: H_NH0673p17
 ----- Summary Statistics -----
 Sequencing vector: M13; 100%
 Sequencing vector: plasmid; 0%
 Chemistry: Dye-terminator Big Dye, 0% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 195107 bases at least Q40
 Consensus quality: 197538 bases at least Q30
 Consensus quality: 198979 bases at least Q20
 Insert size: 199000; agarose-fp
 Insert size: 201728; sum-of-contigs
 Quality coverage: 5.33 in Q20 bases; agarose-fp
 Quality coverage: 5.29 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 12 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 1109: contig of 1109 bp in length
 * 1110 1209: gap of unknown length
 * 1210 5935: contig of 4726 bp in length
 * 5936 6035: gap of unknown length
 * 6036 13020: contig of 6985 bp in length
 * 13021 13120: gap of unknown length
 * 13121 21028: contig of 7908 bp in length
 * 21029 21128: gap of unknown length
 * 21129 27413: contig of 6285 bp in length
 * 27414 27513: gap of unknown length
 * 27514 34551: contig of 7038 bp in length
 * 34552 34651: gap of unknown length
 * 34652 53143: contig of 18492 bp in length
 * 53144 53243: gap of unknown length
 * 53244 77201: contig of 23958 bp in length
 * 77202 77301: gap of unknown length
 * 77302 106156: contig of 28855 bp in length
 * 106157 106257: gap of unknown length
 * 106258 131644: contig of 25388 bp in length
 * 131645 131744: gap of unknown length
 * 131745 158698: contig of 26954 bp in length
 * 158699 158798: gap of unknown length
 * 158799 202828: contig of 44030 bp in length.

FEATURES
 source
 Location/Qualifiers
 1..202828
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RP11-673p17"
 1..1109
 /note="assembly_name:Contig14"
 1210..5935
 /note="assembly_name:Contig18"
 6036..13020
 /note="assembly_name:Contig19"
 13121..21028
 /note="assembly_name:Contig20"
 21129..27413
 /note="assembly_name:Contig21"
 clone_end:17
 vector_side:right"
 27514..34551
 /note="assembly_name:Contig22"
 34652..53143
 /note="assembly_name:Contig23"
 53244..77201
 /note="assembly_name:Contig24"
 77302..106156
 /note="assembly_name:Contig25"
 106257..131644
 /note="assembly_name:Contig26"
 clone_end:SP6
 vector_side:left"
 131745..158698
 /note="assembly_name:Contig27"
 158799..202828
 /note="assembly_name:Contig28"
 BASE COUNT 53011 a 46987 c 47342 g 54379 t 1109 others
 ORIGIN
 Query Match 64.0%; Score 16; DB 72; Length 202828;
 Best Local Similarity 100.0%; Pred. No. 45;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 5 cagatgcagatgacga 20
 Db 46988 CAGGTGCAGGTACGA 47003
 RESULT 43
 AC016186/c
 LOCUS AC016186 245499 bp DNA HTG 12-MAR-2000
 DEFINITION Homo sapiens chromosome 18 clone RP11-8115 map 18, *** SEQUENCING
 IN PROGRESS ***; 70 unordered pieces.
 ACCESSION AC016186
 VERSION AC016186.3 GI:7230177
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 245499)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
 TITLE Unpublished
 JOURNAL 2 (bases 1 to 245499)
 REFERENCE 2 (bases 1 to 245499)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E.,
 TITLE Unpublished
 JOURNAL 2 (bases 1 to 245499)
 REFERENCE 2 (bases 1 to 245499)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
 Baldwin, J., Batta, N., Becker, R., Boguslavsky, L., Boukhalter, B.,
 Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
 Cooke, P., DeBartolomeo, R., Dewar, K., Domingo, M., Donnellan, L., Doyle, M.,
 Ferreira, P., FitzHugh, M., Forrest, C., Funke, R., Gage, D.,
 Galagan, J., Gardy, S., Grant, G., Hagos, B., Heath, A., Horton, L.,
 Howland, J., Johnson, R., Jones, C., Kann, L., Karas, A., Klein, J.,
 Loeck, J., Liu, C., Locke, K., MacDonald, P., Marquis, N.,
 McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Melidim, J.,

TITLE
JOURNAL
COMMENT

Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Sever, P.,
Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testa, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W.-J., Zimmer, A. and Zody, M.
Direct Submission

Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced g1:6649364.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WtBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L4258

Center clone name: 8_I_15

* NOTE: This is a 'working draft' sequence. It currently
* consists of 70 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1174: contig of 1174 bp in length
* 1175 1274: gap of 100 bp
* 1275 2609: contig of 1335 bp in length
* 2610 2709: gap of 100 bp
* 2710 3931: contig of 1222 bp in length
* 3932 4031: gap of 100 bp
* 4032 5121: contig of 1090 bp in length
* 5122 5221: gap of 100 bp
* 5222 6287: contig of 1066 bp in length
* 6288 6387: gap of 100 bp
* 6388 7686: contig of 1299 bp in length
* 7687 7786: gap of 100 bp
* 7787 8797: contig of 1011 bp in length
* 8798 8897: gap of 100 bp
* 8898 10536: contig of 1639 bp in length
* 10537 10636: gap of 100 bp
* 10637 11858: contig of 1222 bp in length
* 11859 11958: gap of 100 bp
* 11959 13375: contig of 1417 bp in length
* 13376 13475: gap of 100 bp
* 13476 15102: contig of 1627 bp in length
* 15103 15202: gap of 100 bp
* 15203 16445: contig of 1243 bp in length
* 16446 16545: gap of 100 bp
* 16546 17637: contig of 1092 bp in length
* 17638 17737: gap of 100 bp
* 17738 19309: contig of 1572 bp in length
* 19310 19409: gap of 100 bp
* 19410 21532: contig of 2123 bp in length
* 21533 21632: gap of 100 bp
* 21633 23166: contig of 1534 bp in length
* 23167 23266: gap of 100 bp
* 23267 25262: contig of 1996 bp in length
* 25263 25362: gap of 100 bp
* 25363 27315: contig of 1953 bp in length
* 27316 27415: gap of 100 bp
* 27416 28722: contig of 1307 bp in length
* 28723 28822: gap of 100 bp
* 28823 30596: contig of 1774 bp in length
* 30597 30696: gap of 100 bp
* 30697 33203: contig of 2507 bp in length
* 33204 33303: gap of 100 bp
* 33304 35408: contig of 2105 bp in length
* 35409 35508: gap of 100 bp

35509 37441: contig of 1933 bp in length
* 37442 37541: gap of 100 bp
* 37542 40401: contig of 2860 bp in length
* 40402 40501: gap of 100 bp
* 40502 42052: contig of 1551 bp in length
* 42053 42152: gap of 100 bp
* 42153 44294: contig of 2142 bp in length
* 44295 44394: gap of 100 bp
* 44395 46328: contig of 1934 bp in length
* 46329 46428: gap of 100 bp
* 46429 48649: contig of 2221 bp in length
* 48650 48749: gap of 100 bp
* 48750 51037: contig of 2288 bp in length
* 51038 51137: gap of 100 bp
* 51138 53348: contig of 2211 bp in length
* 53349 53448: gap of 100 bp
* 53449 55337: contig of 1869 bp in length
* 55338 55437: gap of 100 bp
* 55438 57028: contig of 1591 bp in length
* 57029 57128: gap of 100 bp
* 57129 59830: contig of 2702 bp in length
* 59831 59930: gap of 100 bp
* 59931 62172: contig of 2242 bp in length
* 62173 62272: gap of 100 bp
* 62273 64940: contig of 2668 bp in length
* 64941 65040: gap of 100 bp
* 65041 67862: contig of 2822 bp in length
* 67863 67962: gap of 100 bp
* 67963 70236: contig of 2274 bp in length
* 70237 70336: gap of 100 bp
* 70337 72533: contig of 2197 bp in length
* 72534 72633: gap of 100 bp
* 72634 74796: contig of 2163 bp in length
* 74797 74896: gap of 100 bp
* 74897 77899: contig of 3003 bp in length
* 77900 77999: gap of 100 bp
* 78000 81075: contig of 3076 bp in length
* 81076 81175: gap of 100 bp
* 81175 84346: contig of 3171 bp in length
* 84347 84446: gap of 100 bp
* 84447 87026: contig of 2580 bp in length
* 87027 87126: gap of 100 bp
* 87127 90859: contig of 3733 bp in length
* 90860 90959: gap of 100 bp
* 90960 93599: contig of 2640 bp in length
* 93600 93699: gap of 100 bp
* 93700 96946: contig of 3247 bp in length
* 96947 97046: gap of 100 bp
* 97047 100076: contig of 3030 bp in length
* 100077 100176: gap of 100 bp
* 100177 105333: contig of 5157 bp in length
* 105334 105433: gap of 100 bp
* 105434 109169: contig of 3736 bp in length
* 109170 109269: gap of 100 bp
* 109270 113940: contig of 4671 bp in length
* 113941 114040: gap of 100 bp
* 114041 119295: contig of 5255 bp in length
* 119296 119395: gap of 100 bp
* 119396 120878: contig of 1483 bp in length
* 120879 120978: gap of 100 bp
* 120979 123579: contig of 2601 bp in length
* 123580 123679: gap of 100 bp
* 123680 128211: contig of 4532 bp in length
* 128212 128311: gap of 100 bp
* 128312 132536: contig of 4222 bp in length
* 132537 132636: gap of 100 bp
* 132637 137226: contig of 4590 bp in length
* 137227 137326: gap of 100 bp
* 137327 142521: contig of 5195 bp in length
* 142522 142621: gap of 100 bp
* 142622 147070: contig of 4449 bp in length
* 147071 147170: gap of 100 bp
* 147171 151866: contig of 4696 bp in length

```

* 151867 151966: gap of 100 bp
* 151967 156914: contig of 4948 bp in length
* 156915 157014: gap of 100 bp
* 157015 162849: contig of 5835 bp in length
* 162850 162949: gap of 100 bp
* 162950 169655: contig of 6706 bp in length
* 169656 169755: gap of 100 bp
* 169756 174040: contig of 4285 bp in length
* 174041 174140: gap of 100 bp
* 174141 180979: contig of 6839 bp in length
* 180980 181079: gap of 100 bp
* 181080 190751: contig of 9672 bp in length
* 190752 190851: gap of 100 bp
* 190852 202044: contig of 11193 bp in length
* 202045 202144: gap of 100 bp
* 202145 211327: contig of 9183 bp in length
* 211328 211427: gap of 100 bp
* 211428 224152: contig of 12725 bp in length
* 224153 224252: gap of 100 bp
* 224253 234236: contig of 9984 bp in length
* 234237 245499: contig of 11163 bp in length.
* 245499 245499: contig of 11163 bp in length.

FEATURES
source
1. 245499
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="18"
   /map="18"
   /clone="RP11-8115"
   /clone_11b="RPC1-11 Human Male BAC"

Query Match      64.0%: Score 16; DB 64; Length 245499;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5  caagtcagtcagca 20
         |||||
Db 109551 CAGGTGACGTCAGCA 109536

RESULT 44
AP002552      262278 bp  DNA      BCF      07-MAR-2001
DEFINITION   Escherichia coli O157:H7 DNA, complete genome, section 3/20.
ACCESSION   AP002552 BA000007
VERSION      AP002552.1 GI:13359995
KEYWORDS
SOURCE
ORGANISM
Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA.
Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (sites)
Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C. and Shinagawa,H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
JOURNAL
MEDLINE
REFERENCE
2 (sites)
Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
Hayashi,T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MGI655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
JOURNAL
MEDLINE
REFERENCE
3 (sites)
Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
Shinagawa,H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
JOURNAL
MEDLINE
REFERENCE
4 (sites)
Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C.G., Ohtsubo,E., Nakayama,K., Murata,T.,
Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
Shinagawa,H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
JOURNAL
MEDLINE
REFERENCE
5 (bases 1 to 262278)
Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
Hayashi,T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
Fax:81-6-6879-2047)
genome project.

FEATURES
source
1. 262278
   /organism="Escherichia coli O157:H7"
   /strain="O157:H7"
   /sub_strain="RIMD 0509952"
   /db_xref="taxon:83334"
   /gene="ECs0538"
   /note="similar to YBAS_ECOLI g111786693 percent identity
98 in 310 aa (Conserved in E.coli K-12)"
   /codon_start=1
   /transl_table=1
   /evidence="not_experimental"
   /product="putative glutaminase"
   /protein_id="BAB33961.1"
   /db_xref="GI:13359995"
   /translation="MUDANKLQQAQAVDQAYTQFHSLSNGQNDYIPIIAPNPGQLAAYA
IVTSDGNYVSAGSDSDYRFALESISKVCTLALALBDYQPAVDKIGADPTGLPNSVI
ALETHGKPLSPLVNAGAIAVTSILINENAEQRORILIIIOQLAGELVALSDVNS
EOTTFHNRATAMILLYSAGLYCDAMEACDVTYTRQCSITLIVTELATIGATLAAGVY
PLTHRRVADQNVPIYILAEKMMEGLYGSGGMAYRVGLPKSGVGGGILAVVPGVMGI
AAFPSPDDEGNSVRGQKMAVSVAKQIGYNFKG"
   /gene="ECs0538"
   /note="similar to YBAS_ECOLI g111786693 percent identity
98 in 310 aa (Conserved in E.coli K-12)"
   /codon_start=1
   /transl_table=1
   /evidence="not_experimental"
   /product="putative amino acid/amino transport protein"
   /protein_id="BAB33962.1"
   /db_xref="GI:13359997"
   /translation="MMNREGNGNGKPKLIGNVVSGIGAMYGAGIFALIGOAALIMEA
STWAFARCGIVAMFSGYATYARLCAISPSNGGIIIDFRRGIGANVPSIALSLYLTL
AVSIAMVARAFCAIVADPLHEGSOEHEHILLYLAGITLAVMLTFNSLSNHAGREVLV
VGIKMMLILLIAGVMSILOPAHISVSPSSGAFSCIGITFLAYAGFGMAAARAK
VKDPOVIMRAFIIVAGVITLLYISLAILVLSDVSALELEKADTAQAQASPLIIVY
GVYIVIGALLATASAINANLFAVENIMDNNGSRELPKRNKSLMDSOTMGNTIIVY
LIMLTAINLGSILASVASATFELICYLAFPVATIRLPHDIHASPILIIIVGIVMLIYI
VGFITSLMSQSRALIMITGAILLSLIVAMMRNKTY"
   2724..4079

```

gene


```

CDS
/gene="ECs0540"
/2724..4079
/feature="putative transporter (outer membrane protein), similar to outer membrane transport proteins e.g. Cyab protein [Bordetella pertussis] gi1177991|sp|p10921|CYAB_BORPE percent identity 25 in 385 aa"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="putative outer membrane transport protein"
/protein_id="BAB33963.1"
/db_xref="GI:13359998"
/translation="MARFORKNRKNKGLIFPISPMWGEAIAPIPLPOMANAAPVTPV
AQLQESILRAEARNPVTQAQIGIGEAQIDEAKSAMYPHVGTLGNAGSPKQDTS
SGRLDNVSYGILTQLVDFGKTNNDINQTAARDYRFLMATLTVDAVEKTAIAM
EVRVQALCDAAQAHINSLERNYMMALRANAGLNSIDELQSIPLVIAENLRQSAQY
AOMASAKAQLAVLTGVOPEAIAAPPALADQPSLKIDQSIPLVIAENLRQSAQY
QVETKRAQWPTISIOGKTRVOTSDPSYDDOQLNVNAPLYGGAGVSAQOYAEQ
OKISASOVEQAKLDVLRASVATNMTGARGREAGLAQSESAKTRDVTONEKRLAK
KSLNDLITVEQDFVQAQSAETINAMTDGVAANYAAVNNLIPLAGIKQGLNDLPDL
K"
4182..8567
/gene="ECs0541"
/4182..8567
/feature="unknown, similar to hypothetical proteins e.g. [Synechocystis sp. strain PCC 6803] gi17469433|pif|S76109 percent identity 33 in 1384 aa, also similar to RTX protein [Aeromonas salmonicida] gi16752871|gb|AAE27914.1|AAE218037_1 percent identity 33 in 138"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="hypothetical protein"
/protein_id="BAB33964.1"
/db_xref="GI:13359999"
/translation="MSLITDVISRKTSVKQTLINPGDVTYVIEPSVYVQAQASAVA
RYVREGNDLILYMDQGVIRKNGYFLEAANAEQSELVFADGQQLHITPADAGGL
APVELTQATIESIAPFLDTVAQTSAPFWMLGAAGVGGALCALIAGSDGSKTE
VTNPTPEAPGNATPSPIVTDNQGQRLATDIDPTPPSGSGGQAGATQID
SNGNTASTOVNDNHHMSVLPQSAGEHMSVVOYVITTTAGSTTTITAGNSQAV
OVATTAGDNITINASEQAQFTLSTSSHLQGLTVTLNKGTYTTSVANGASAVOV
PTADQAQLEGNOAVLVSGKDATNGTVAQLLVDPQPLTAINTAQDNITISAAB
NVALVSGTSMNAEGQTVTLNKGSHATVAGSDGTPLPALEYOALAEQNAVANA
SVSDRAGNTSHSNFTVDTSAVYAVNTVAGDILNNAEQAQIISGOVSGASPCD
TVYKIGTHVIGIYIADGSNNVALDPAVYPTLDRGANTLEVTYDAGAGMGASRAI
TLVGVSLITNTVSGDDITISAEKGAFLITGSTQAETGQVTVTLAQSFTTYYQ
ADGSMSLTPPAAMGNLPDGAVALTASVTDLSGNTSRTITVDSQAPALSIDPLA
DNIINAAESGQDLEPTGTDAQPGQTVTLNGYGVVQPDGTWSPVPAANVGL
ADAGATVYASVNDVAGNNSVRYALVADATPPVYINPVATDNIINPEHAQALISG
TVTGAQAGDITVTLNNVDYTTVYVAGSNNLSGPASVSGLAGSVSVSYVDK
NTGOSLITVYNTAPILIGINSIAGDVIYNSAEKADLOITGNSDQVNIATYTLNG
QNTYTTPDASGMSVTPASAVTALGQAVTVAATVSDTNGNSTASHNIVYSLALG
VTINPVATDILINAAEGVQVOTISGOVYAGDGVTVITLGGNTATVGSNLVSLALG
VPADIALGNGDLTVNASVTNONGNCSGRITLIDANLPGLVNDVAGDVNIIIE
HGOALVYGGSSGLAESTPLVTIINNEYTVAQDSWGSVGAADVAAVSPATVNI
AVSGSSAGNSVSIITHPVYDLTPAATINTIADVDYINAEKACADLTLSSTNNVP
GQTVYFPGCKNTYASVADGSMATVPAADLALPSSASALASVNSINGNSAVH
NYSVDSAPITITINTVASDNIYNASEADAGTVSGSTTAEQGVITTLNTPYQVTO
ATVADGSMSTINIPADLEALITDGSHTLITATNKGAPASTHNLAVDLTPVLITN
TIAADDIINATEHGOALVIGSSITGEGGADVTVTLNLSKYTTTLTLDASGMSVGPAA
DVTALGSPQTVTATVTDACNSDN"
8775..24650
/gene="ECs0542"
/8775..24650
/feature="unknown, similar to hypothetical proteins e.g. [Synechocystis sp. strain PCC 6803] gi17469433|pif|S76109 percent identity 36 in 2014 aa, similar to RTX protein [Aeromonas salmonicida]

```

```

Query Match      64.0% Score 16: DB 2: Length 262278;
Best Local Similarity 100.0%: Pred. No. 43;
Matches 16: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9  tgcagtcagcagcgt 24
          |||||
Db 152470 TGCAGTGCAGCAGCCTT 152485

RESULT 45
HSMX1B
LOCUS      HSMX1B      313064 bp      DNA      PRI      23-SEP-2000
DEFINITION Homo sapiens chromosome 21 from 5 PACs and 5 Cosmids map
                21q22.2, D21S349-MX1; segment 2/2, complete sequence.
ACCESSION   AL442167 AJ011929
VERSION     AL442167.1 GI:10303260
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 313064)
AUTHORS     Ramser, J., Francis, F., Beck, A., Hennig, S., Klages, S., Borzym, K.,
             Langer, I., Steffens, C., Hildmann, T., Dagand, E., Yaspo, M.,
             Reinhardt, R. and Lehrach, H.
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 313064)
AUTHORS     MPMG.
TITLE       Direct Submision
JOURNAL     Submitted (01-OCT-1998) MPMG, Abt. Lehrach, Max Planck Institut
             fuer Molekulare Genetik, Innesstrasse 73, Berlin, 14195, Germany
COMMENT     Clones received from Resource Centre of the Human Genome Project at
             the Max-Planck-Institut for Molecular Genetics.
             232889J..294336 Sequence from clone K8447A5, accession no. AP001609
             (DBJ), sequenced at KEIO university, Tokyo, Japan.
             Bases 1..50 overlap with AL442166.
FEATURES             location/Qualifiers
             1..313064
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="21"
             /map="21q22.2, D21S349-MX1"
             /clone="PAC RPCI-1 247E2"
             /clone="PAC RPCI-1 146B4"
             /clone="PAC RPCI-1 141D16"
             /clone="PAC RPCI-1 269A14"
             /clone="PAC RPCI-1 265B9"
             /clone="cosmid LINC116 44C5"
             /clone="cosmid LINC116 16H18"

```

source
/clone="cosmid L1NLC116 14C10"
/clone="cosmid L1NLC116 25D2"
/clone="cosmid L1NLC116 87D5"
/clone.lib="L121NC02, Chromosome 21 specific cosmid
library, Lawrence Livermore National Laboratory (LLNL),
creator: Pieter de Jong; PAC: RPI1,3-5, Roswell Park
Cancer Institute, creator: Pieter de Jong, P.Ioannou"
29289, .294336
/note="sequence from clone KB47A5, accession no. AP001609
(DDBJ), sequenced at KEIO university, Tokyo, Japan."
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"

BASE COUNT 83213 a 71272 c 73589 g 84990 t
ORIGIN

Query Match 64.0%: Score 16; DB 93; Length 313064;
Best Local Similarity 100.0%: Pred. No. 41;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 gtcaggtcaggtcag 18
|||||
Db 182813 CTCAGGTGTCAGTCAG 182828

Search completed: October 9, 2001, 15:49:28
Job time: 14954 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:53:33 ; Search time 469.84 Seconds
(without alignments)
33.410 Million cell updates/sec

Title:	US-09-396-196F-10
Perfect score:	25
Sequence:	1 tcgtcaggtgcaggtcagcacglttg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

word size : 9

Total number of hits satisfying chosen parameters: 34718

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Listing first 45 summaries

```

Database :
N.GeneSeq_0601.*
1: /SIDS1/gcgdata/geneSeq/geneSeqc/NAl1980.DAT.*
2: /SIDS1/gcgdata/geneSeq/geneSeqc/NAl1981.DAT.*
3: /SIDS1/gcgdata/geneSeq/geneSeqc/NAl1982.DAT.*
4: /SIDS1/gcgdata/geneSeq/geneSeqc/NAl1983.DAT.*
5: /SIDS1/gcgdata/geneSeq/geneSeqc/NAl1984.DAT.*
6: /SIDS1/gcgdata/geneSeq/geneSeqc/NAl1985.DAT.*
7: /SIDS1/gcgdata/geneSeq/geneSeqc/NAl1986.DAT.*
8: /SIDS1/gcgdata/geneSeq/geneSeqc/NAl1987.DAT.*
9: /SIDS1/gcgdata/geneSeq/geneSeqc/NAl1988.DAT.*
10: /SIDS1/gcgdata/geneSeq/geneSeqc/NAl1989.DAT.*
11: /SIDS1/gcgdata/geneSeq/geneSeqc/NAl1990.DAT.*
12: /SIDS1/gcgdata/geneSeq/geneSeqc/NAl1991.DAT.*
13: /SIDS1/gcgdata/geneSeq/geneSeqc/NAl1992.DAT.*
14: /SIDS1/gcgdata/geneSeq/geneSeqc/NAl1993.DAT.*
15: /SIDS1/gcgdata/geneSeq/geneSeqc/NAl1994.DAT.*
16: /SIDS1/gcgdata/geneSeq/geneSeqc/NAl1995.DAT.*
17: /SIDS1/gcgdata/geneSeq/geneSeqc/NAl1996.DAT.*
18: /SIDS1/gcgdata/geneSeq/geneSeqc/NAl1997.DAT.*
19: /SIDS1/gcgdata/geneSeq/geneSeqc/NAl1998.DAT.*
20: /SIDS1/gcgdata/geneSeq/geneSeqc/NAl1999.DAT.*
21: /SIDS1/gcgdata/geneSeq/geneSeqc/NAl2000.DAT.*
22: /SIDS1/gcgdata/geneSeq/geneSeqc/NAl2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	25	100.0	1041	20	AAV01303	E. coli biotin synthase
2	25	100.0	1084	10	AAV01329	E. coli B gene
3	25	100.0	1121	7	AAV0496	Sequence encoding
4	25	100.0	5872	15	AAV02386	Biotin-biosynthesis
5	25	60.0	353	21	AAV05641	Eucalyptus grandis
6	15	60.0	630	16	AAV78948	Human immunoglobulin
7	15	60.0	4674	20	AAV90421	Human ataxin-2
8	14	56.0	269	21	AAV31747	Plant microsatellite
9	14	56.0	327	21	AAV00048	Human secreted protein
10	14	56.0	374	22	AAV64981	Human secreted protein
11	14	56.0	423	14	AAV05597	Sequence encoding

C	12	14	56.0	423	20	AAZ39427
C	13	14	56.0	423	20	AAZ233965
C	14	14	56.0	444	18	AA7B0181
C	15	14	56.0	447	20	AAZ20404
C	16	14	56.0	437	16	AAO78943
C	17	14	56.0	981	21	AAAI5905
C	18	14	56.0	1100	21	AAAI59195
C	19	14	56.0	1130	21	AAA39064
C	20	14	56.0	1135	21	AAA39087
C	21	14	56.0	1161	22	AAA91019
C	22	14	56.0	1173	21	AAZ981136
C	23	14	56.0	1186	21	AAAC8064
C	24	14	56.0	1191	19	AA7B987266
C	25	14	56.0	1366	22	AAAF4623
C	26	14	56.0	1634	19	AAZ66282
C	27	14	56.0	1634	19	AAV37360
C	28	14	56.0	37856	21	AAAI1350
C	29	14	56.0	534720	19	AAV30458
C	30	14	56.0	536165	19	AAV30458
C	31	13	52.0	18	12	AAO13667
C	32	13	52.0	123	20	AAH85925
C	33	13	52.0	225	14	AAH85925
C	34	13	52.0	276	21	AAO60177
C	35	13	52.0	363	20	AAAC00049
C	36	13	52.0	363	20	AAH85925
C	37	13	52.0	375	21	AAZ95272
C	38	13	52.0	378	18	AAZ22004
C	39	13	52.0	378	18	AA7C1237
C	40	13	52.0	400	21	AA7C1237
C	41	13	52.0	423	20	AAAC00033
C	42	13	52.0	437	21	AAZ92070
C	43	13	52.0	472	21	AAZ42885
C	44	13	52.0	477	20	AAZ5419
C	45	13	52.0	527	20	AAZ20408
						AAV87894

ALIGNMENTS

RESULT	1
AAAX01303	
ID	AAAX01303 standard; DNA; 1041 BP.
XX	
AC	AAAX01303;
XX	
DT	12-APR-1999 (first entry)
XX	
DE	E. coli biotin synthetase (BioB) coding sequence.
KW	DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW	biotin synthase; biotin production; vitamin H; BioB; ss.
XX	
OS	Escherichia coli.
XX	
PN	US5869719-A.
XX	
PD	09-FEB-1999.
XX	
PF	30-APR-1997; 97US-0846338.
XX	
PR	30-APR-1997; 97US-0846338.
PR	08-MAR-1995; 95US-0401068.
XX	
PA	(NOVS) NOVARTIS FINANCE CORP.
XX	
PI	Patton DA;
XX	
DR	WPI; 1999-152902/13.
DR	P-PSDB; AAW73906.
XX	
PT	Transgenic plants with high biotin levels - transformed with DNA
PT	encoding di-amino-pelargonic acid amino-transferase or biotin
PT	synthase

XX Example 2: Column 37-40; 34pp: English.

CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC be used in the transgenic plant of the invention. The transgenic plant,
CC plant cell or plant tissue is transformed with a chimeric gene encoding
CC diaminopelargonic acid (DAP) aminotransferase or biotin synthetase and
CC produces more biotin than a non-transgenic plant, cell or tissue. The
CC plant is used as an improved dietary source of biotin (vitamin H) for
CC humans or animals.

CC Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.00025;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tcgtcaggtgcaggtcagcagcttg 25
Db 111 tcgtcaggtgcaggtcagcagcttg 135

RESULT 2

AA062386 standard; DNA; 1084 BP.

AA062386;

15-FEB-1990 (first entry)

E.coli Bio B gene.

E.coli Bio B gene; biotin.

Escherichia coli.

Key Location/Qualifiers

FT CDS 24..1064

FT CDS /*tag=a

GB2216530-A.

11-OCV-1989.

17-MAR-1989; 89GB-0006210.

22-MAR-1988; 88GB-0006804.

17-MAR-1989; 89GB-0006210.

(UKAG-) UK MIN. AGRIC. FISH.

Pearson BM, McKee RA;

WPI; 1989-295085/41. P-PSDB P91392

Plasmid contg. gene(s) for expression of biotin synthetase enzymes

derived from E.coli and capable of replication and expression in other

microorganisms, esp. yeast.

Table 3; page 33-4; 52pp: English.

The gene can be used in a plasmid for expression of enzymes of the biotin

synthetic pathway. Pref. control sequences for expression in S.cerevisiae

are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for

Lactobacillus. Insertion of bio B improves biotin yields in

microorganisms which export biotin, or enables growth in media contg.

little or no biotin of organisms unable to synthesise biotin for their

Query Match 100.0%; Score 25; DB 10; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.00025;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tcgtcaggtgcaggtcagcagcttg 25
Db 134 tcgtcaggtgcaggtcagcagcttg 158

RESULT 3

AA060496 standard; DNA; 1121 BP.

AA060496;

17-OCT-1991 (first entry)

Sequence encoding biotin synthesising enzyme.

Biotin synthetic enzyme; E.coli; desthiobiotin; ds.

Key Location/Qualifiers

FT CDS 42..1082

FT CDS /*tag=a

JP61149091-A.

07-JUL-1986.

24-DEC-1984; 84JP-0272605.

24-DEC-1984; 84JP-0272605.

(NIPS) NIPPON SODA KK.

WPI; 1986-216622/33.

P-PSDB; AAP60536.

Double stranded DNA encoding biotin synthesising enzyme -

comprises transformed mutant E.coli strain contg. cyclic doubled

stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.

Disclosure; Page 534; 23pp; Japanese.

The sequence may be expressed by a transformed E.coli host, cultured

in a medium containing desthiobiotin.

Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;

Best Local Similarity 100.0%; Pred. No. 0.00025;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 tcgtcaggtgcaggtcagcagcttg 25

Db 152 tcgtcaggtgcaggtcagcagcttg 176

AA062386 standard; DNA; 5672 BP.

AA062386;

16-NOV-1994 (first entry)

Biotin-biosynthesis genes contg. plasmid PB030A-15/9.

Biotin; expression; enterobacteria; vitamin H; synthesis;

plasmid; PB030A-15/9; bioB; bioC; bioD; bioA;

promoter plac; biotin synthetase; KAPA synthetase;

8-amino-7-oxononanoate synthetase; pimeloyl-CoA; DNB synthetase;

KW dethiobiotin synthase; DAPA synthase;
 KW S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
 KW seborrhoea; dermatitis; ds.
 XX
 OS Escherichia coli DSM498.
 XX
 FH
 FH Key Location/Qualifiers
 FT 1..96
 FT promoter
 FT /tag= a
 FT /function= "promoter plac"
 FT /evidence= EXPERIMENTAL
 FT 23..28
 FT -35_signal
 FT /tag= b
 FT /standard_name= "promoter plac"
 FT 45..50
 FT -10_signal
 FT /tag= c
 FT /evidence= EXPERIMENTAL
 FT /standard_name= "promoter plac"
 FT 105..109
 FT RBS
 FT /tag= d
 FT /evidence= EXPERIMENTAL
 FT /standard_name= "bioB RBS no. 9"
 FT 117..1157
 FT CDS
 FT /tag= e
 FT /product= "biotin synthase"
 FT /evidence= EXPERIMENTAL
 FT /gene= "bioB"
 FT /number= 1
 FT 1141..1146
 FT RBS
 FT /tag= f
 FT /standard_name= "bioF RBS"
 FT 1154..2311
 FT /tag= g
 FT /EC_number= 2.3.1.47
 FT /product= "KAPA synthase"
 FT /evidence= EXPERIMENTAL
 FT /gene= "bioF"
 FT /number= 2
 FT /standard_name= "8-amino-7-oxononanoate synthase"
 FT 2284..2288
 FT RBS
 FT /tag= h
 FT /standard_name= "bioC RBS"
 FT 2295..3050
 FT CDS
 FT /tag= i
 FT /function= "involved in pimeloyl-CoA synthesis"
 FT /product= "protein"
 FT /gene= "bioC"
 FT /number= 3
 FT 3030..3033
 FT RBS
 FT /tag= j
 FT /standard_name= "bioD RBS"
 FT 3043..3753
 FT CDS
 FT /tag= k
 FT /EC_number= 6.3.3.3
 FT /product= "DPA synthase"
 FT /evidence= EXPERIMENTAL
 FT /gene= "bioA"
 FT /number= 5
 FT /standard_name= "S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase"
 FT 3712..3750
 FT misc_RNA
 FT /tag= l
 FT /note= "bioid15 substitution"
 FT 3742..3746
 FT RBS
 FT /tag= m
 FT /standard_name= "bioA RBS"
 FT 3750..5039
 FT CDS
 FT /tag= n
 FT /EC_number= 2.6.1.62
 FT /product= "DAPA synthase"
 FT /evidence= EXPERIMENTAL
 FT /gene= "bioA"
 FT /number= 5
 FT /standard_name= "S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase"

FT RBS 5088..5093
 FT /tag= o
 FT /standard_name= "ORFI RBS"
 FT 5098..5574
 FT CDS
 FT /tag= p
 FT /function= "unknown, involved in biotin synthesis"
 FT /product= "protein"
 FT /evidence= EXPERIMENTAL
 FT /gene= "ORFI"
 FT /number= 6
 FT 5583..5644
 FT terminator
 FT /tag= q
 FT /standard_name= "rho-independent transcriptional terminator"
 FT 5583..5605
 FT stem_loop
 FT /tag= r
 FT 5583..5605
 FT W09408023-A.
 FT 14-APR-1994.
 FT PD
 FT 01-OCT-1993; 93WO-EP02688.
 FT PF
 FT 02-OCT-1992; 92CH-0003124.
 FT PR 15-JUL-1993; 93CH-0002134.
 FT XX
 FT (LONZ) LONZA AG.
 FT PA
 FT Birch O, Brass J, Fuhrmann M, Shaw N;
 FT PI
 FT WPI: 1994-135587/16.
 FT DR P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
 FT XX
 FT Biotechnological biotin prodn. using enterobacterial biotin gene
 FT PT - providing vitamin H in high yield
 FT PS
 FT Claim 1; Fig 6, Page 47-55 and 60-65; 92pp: German.
 FT XX
 FT The sequence is derived from plasmid pB030A-15/9 contg. the
 FT CC bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
 FT CC of biotin, arranged in a transcription unit. Microorganisms
 FT CC contg. these DNA fragments or plasmids may be used in the prodn.
 FT CC of biotin. Biotin (vitamin H) may prevent seborrhoea, dermatitis,
 FT CC loss of appetite and tiredness.
 FT XX
 FT Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;
 FT SQ
 FT Query Match 100.0%; Score 25; DB 15; Length 5872;
 FT Best local Similarity 100.0%; Pred. No. 0.00024;
 FT Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 FT
 FT QY 1 tcgtcaggtcaggtcaggtcaggtg 25
 FT |||||||||||||||||||||||||
 FT Db 227 tcgtcaggtcaggtcaggtcaggtg 251
 FT
 FT RESULT 5
 FT AAC56641/c
 FT ID AAC56641 standard; DNA: 353 BP.
 FT XX
 FT AAC56641;
 FT AC
 FT XX
 FT 25-JUN-2001 (first entry)
 FT DT
 FT XX
 FT Eucalyptus grandis transcription factor DNA sequence #512.
 FT DE
 FT XX
 FT Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
 FT KW poplar; sweetgum; leaf; mahogany; bZIP; G-box binding factor;
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox;
 KW homeodomain zipper; LIM domain; AP2; ERBB; zinc finger domain;
 KW type 2 Cys2His2; COAT box element; MTR; ss.

```

XX OS Eucalyptus grandis.
XX PN WO200053724-A2.
XX PD 14-SEP-2000.
XX PF 09-MAR-2000; 2000WO-US06112.
XX PR 11-MAR-1999; 99US-0266513.
XX PR 18-AUG-1999; 99US-0149485.
XX PA (GENE-) GENESTIS RES & DEV CORP LTD.
XX PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX PI Wood M, McGrath A, Shenk MA, Glenn M;
XX DR WPI: 2000-579369/54.
XX PT New isolated polynucleotide encoding a plant transcription factor for
XX PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
XX PT having modified gene expression or modified activity of a polypeptide
XX PS
XX PS Claim 1; Page 482; 747pp; English.
XX CC The present invention relates to novel plant transcription factors from
XX CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
XX CC sequence for one such transcription factor. The transcription factor may
XX CC be used to produce a plant having modified gene expression such as a
XX CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
XX CC mahogany species or to modify the activity of a polypeptide in a plant.
XX CC The transcription factors of the present invention are members from the
XX CC following families of regulatory proteins: bZIP, bZIP family of G-box
XX CC binding factors, basic helix-loop-helix zipper, bZIP family of G-box
XX CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
XX CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
XX CC
XX SO Sequence 353 BP; 117 A; 59 C; 98 G; 78 T; 1 other;
OY 1 tcgtcaggtcaggt 15
    |||
Db 326 TCGTCAGTCACGCT 312
Query Match 60.0%; Score 15; DB 21; Length 353;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 tcgtcaggtcaggt 15
    |||
Db 326 TCGTCAGTCACGCT 312
RESULT 6
AAO78948/C
ID AAO78948 standard; DNA: 630 BP.
XX AC
XX AC AAO78948;
XX DT 01-AUG-1995 (first entry)
XX DE Human immunoglobulin Vh gene #10.
XX KW Primer; PCR; amplify; human; immunoglobulin; variable; heavy chain;
XX KW cosmid; placenta; vector; PUB81; E.coli; mammalian; ds.
XX OS Homo sapiens.
XX FH
XX FH Key Location/Qualifiers
XX FT 71..495
XX FT /*tag= a
XX FT /product= human immunoglobulin variable heavy chain
XX FT intron 114..199
XX FT /*tag= b
XX FT misc_signal 322..324

```

```

FT FT /*tag= c
FT FT /transl_except= unused termination codon
FT FT misc_signal 373..375
FT FT /*tag= d
FT FT /transl_except= unused termination codon
FT FT misc_signal 496..498
FT FT /*tag= e
FT FT /note= "miscellaneous signal, does not conform to
FT FT termination or splice site sequence"
XX PN WO9426895-A.
XX PD 24-NOV-1994.
XX PF 10-MAY-1993; 93WO-JP00603.
XX PF 10-MAY-1993; 93WO-JP00603.
XX PR 10-MAY-1993; 93WO-JP00603.
XX PA (NISB ) JAPAN TOBACCO INC.
XX PI Honjo T, Matsuda F;
XX PI WPI: 1995-006791/01.
XX DR P-PSDB; AAR6304.
XX DR
XX PT DNA fragment comprising human immunoglobulin Vh genes - for the
XX PT production of human immunoglobulin in mammalian hosts
XX PS
XX PS Claim 19; Page 43-44; 130pp; Japanese.
XX CC A series of genes (AAO78939-79002) encoding human immunoglobulin
XX CC variable heavy chains. The genes were isolated and cloned from a series
XX CC of cosmid constructs: Y202; Y103; Y21; Y6; Y24; 3-31; M84; M18 and M31,
XX CC by PCR amplification using primers AAO78917-38. The genes are subdivided
XX CC into 5 families of Vh genes. The fragments cover a region of 800 kb.
XX CC The DNA fragments were isolated from high molecular weight DNA from
XX CC human placenta. The DNA was partially digested with TagI restriction
XX CC enzyme. The fragments were separated by gel electrophoresis and 35-45 kb
XX CC fragments were collected. The fragments were ligated with ClaI-digested
XX CC cosmid vector PUB81. The ligation products were in vitro packed and
XX CC infected into E.coli 490A. The fragments were then subcloned by colony
XX CC hybridisation. The Vh genes and the DNA fragments encoding them are
XX CC useful in producing human immunoglobulin in mammalian hosts.
XX SO Sequence 630 BP; 141 A; 179 C; 148 G; 162 T; 0 other;
OY 6 aggtcaggtcagca 20
    |||
Db 280 AGGTCAGTCACGCA 266
Query Match 60.0%; Score 15; DB 16; Length 630;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 6 aggtcaggtcagca 20
    |||
Db 280 AGGTCAGTCACGCA 266
RESULT 7
AAO90421/C
ID AAO90421 standard; CDNA: 4674 BP.
XX AC
XX AC AAO90421;
XX DT 29-SEP-1999 (first entry)
XX DE Human ataxin-2 like gene.
XX DE Human ataxin-2 like gene.
XX KW Human; ataxin-2 like gene; ataxin-2 like protein; ataxia; deafness;
XX KW cardiomyopathy; neurological disease; cancer; AIDS; diagnosis; ss.
XX OS Homo sapiens.
XX FH
XX FH Key Location/Qualifiers
XX FT 63..3218

```

```

FT      /*tag= a
XX      /product= "ataxin-2 like protein"
XX
XX      MO936527-A1.
XX
XX      22-JUL-1999.
XX
XX      19-JAN-1998; 98WO-CN00009.
XX
XX      19-JAN-1998; 98WO-CN00009.
XX
XX      (UYHU-) UNTV HUNAN MEDICAL.
XX
XX      Deng H, Liu C, Wang D, Xia J;
XX      WPI; 1999-458463/38.
XX      P-PSDB; AAT29321.
XX
XX      Ataxin-2 like protein, and related polynucleotides, useful in
XX      treatment and diagnosis of ataxia,
XX
XX      Claim 4; Page 21-23; 34pp; English.
XX
XX      The present sequence represents a human ataxin-2 like gene. Ataxin-2
XX      like polynucleotides (1), vectors containing (1) and recombinant host
XX      cells are useful for recombinant production of Ataxin-2 like protein.
XX      Both (1), ataxin-2 like protein and antibodies against ataxin-2 like
XX      protein are useful as research reagents, for screening assays and in
XX      diagnostic assays. Antagonists and agonists of ataxin-2 like protein
XX      can be used to inhibit or enhance, respectively, the activity of
XX      ataxin-2 like protein or expression of (1). Anti-ataxin-2 like protein
XX      antibodies and ataxin-2 like protein or its fragments can be used
XX      in vaccines. In particular, the proteins, antibodies, agonists and
XX      antagonists can be used for treating, e.g. ataxia, cardiomyopathy,
XX      deafness, neurological disease, cancer and AIDS, related to both an
XX      excess and insufficient amounts of ataxin-2 like.
XX
XX      Sequence 4674 BP; 971 A; 1566 C; 1244 G; 893 T; 0 other;
XX
XX      Query Match      60.0%; Score 15; DB 20; Length 4674;
XX      Best Local Similarity 100.0%; Pred. No. 26;
XX      Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX      QY      5 caggtcaggtcagc 19
XX      |||
XX      Db      3741 CAGGTCCAGTCCAGC 3727
XX
XX      RESULT 8
XX      AAA31747
XX      ID AAA31747 standard; DNA; 269 BP.
XX
XX      AC      AAA31747;
XX
XX      DT      05-JUL-2000 (first entry)
XX
XX      DE      Plant microsatellite marker #708.
XX
XX      KW      Plant microsatellite sequence; core repeat sequence; detection; probe;
XX      DNA polymorphism; genome mapping; physical mapping; fingerprinting;
XX      variety identification; genetic variability evaluation; primer; ss.
XX
XX      OS      Eucahyptus grandis.
XX
XX      PN      WO967421-A1.
XX
XX      PD      29-DEC-1999.
XX
XX      PF      25-JUN-1999; 99WO-NZ00092.
XX
XX      PR      25-JUN-1998; 98US-0105307.
XX

```

```

PA      (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
PA      (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX      PI      Havukkala TJ, Bloksberg LN, Glenn M;
XX      WPI; 2000-116958/10.
XX
XX      DR      New plant microsatellite markers and associated flanking species for
XX      the detection of polymorphic genetic markers -
XX
XX      PS      Claim 1; Page 285; 392pp; English.
XX
XX      Sequences AAA31040-A32093 represent novel plant microsatellite sequences
XX      and associated flanking species. The sequences comprise a central core
XX      repeat sequence, especially selected from the sequences AAA32094-A32096
XX      with left and right flanking sequences. The polynucleotide sequences
XX      can be used in the detection of DNA polymorphisms, in genome mapping,
XX      in physical mapping, in positional cloning of genes, in variety
XX      identification and in evaluation of genetic variability within and
XX      between plant tissues, populations, cultivars, species and species
XX      groups. They may also be used to design hybridization probes for
XX      oligonucleotide fingerprinting and library screening and to design
XX      primers for microsatellite-primed PCR. Microsatellite markers are
XX      useful to locate specific economically useful genes in plant genomes.
XX
XX      SEQ      Sequence 269 BP; 63 A; 62 C; 61 G; 83 T; 0 other;
XX
XX      Query Match      56.0%; Score 14; DB 21; Length 269;
XX      Best Local Similarity 100.0%; Pred. No. 95;
XX      Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX      QY      11 cagtcagcagcgtt 24
XX      |||
XX      Db      76 cagtcagcagcgtt 89
XX
XX      RESULT 9
XX      AAC00048/C
XX      ID AAC00048 standard; cDNA; 327 BP.
XX
XX      AC      AAC00048;
XX
XX      DT      06-OCT-2000 (first entry)
XX
XX      DE      Human secreted protein 5' EST, SEQ ID NO: 46.
XX
XX      KW      Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX      gene therapy; chromosome mapping; ss.
XX
XX      OS      Homo sapiens.
XX
XX      PN      EP1033401-A2.
XX
XX      PD      06-SEP-2000.
XX
XX      PF      21-FEB-2000; 2000EP-0200610.
XX
XX      PR      26-FEB-1999; 99US-0122487.
XX
XX      PA      (GSET ) GENSET.
XX
XX      PI      Dumas Milne Edwards J, Duclert A, Giordano J;
XX      WPI; 2000-500381/45.
XX      P-PSDB; AAG00042.
XX
XX      New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX      obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX      diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX      Claim 1; SEQ ID 46; 71pp + CD-ROM; English.
XX

```

CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.

SQ Sequence 327 BP; 70 A; 102 C; 75 G; 69 T; 11 other;

Query Match 56.0%; Score 14; DB 21; Length 327;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtcaggtcagc 19
 |||||
 DB 144 AGGTGACGTGACG 131

RESULT 10
 AAF64981/C
 ID AAF64981 standard; cDNA; 374 BP.

AC AAF64981;

AC 09-APR-2001 (first entry)

DE Novel human polynucleotide, SEQ ID NO: 737.

XX Human; cytostatic; gene therapy; colon cancer; prostate cancer;

KM breast cancer; lung cancer; cancer detection; ss.

XX Homo sapiens.

XX WO200102568-A2.

PD 11-JAN-2001.

PF 30-JUN-2000; 2000WO-US18374.

XX 02-JUL-1999; 99US-0142310.

PR 02-JUL-1999; 99US-0142311.

PA (CHTR) CHIRON CORP.

XX (HYSE-) HYSEQ INC.

PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;

PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;

PI Kirtenjakov R, Drmanac S, Dickson M, Labat I, Lesnikowicz D;

PI Kita D, Garcia V, Jones LW, Strache-Grain B;

XX WPI: 2001-091805/10.

PT Library of polynucleotides for diagnosing a cancerous state of a
 PT mammalian cell and detecting cancer, particularly of the colon or
 PT prostate, comprises 3351 human polynucleotide sequences -

PS Claim 9; Page 650; 1046pp; English.

XX The present sequence is one of 3351 sequences in a library of human
 CC polynucleotides. The library is used to detect differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell and can
 CC detect colon, prostate, breast and lung cancer. The library can be used
 CC to produce probes for detection of mRNA and to produce additional copies
 CC of the polynucleotides. The probes can be used for chromosome mapping of

CC the polynucleotide and for detection of transcription levels. Ribozymes
 CC or antisense oligonucleotides can be generated. The polynucleotides and
 CC their gene products are used as genetic or biochemical markers (e.g. in
 CC blood or tissues) that will detect the earliest changes along the
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and
 CC preventive interventions. The polynucleotides, polypeptides and
 CC antibodies against them can be used in pharmaceutical compositions to
 CC treat the cancers and proliferative disorders such as neoplasia,
 CC dysplasia and hyperplasia.

SQ Sequence 374 BP; 71 A; 116 C; 118 G; 68 T; 1 other;

Query Match 56.0%; Score 14; DB 22; Length 374;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtcaggtcagc 19
 |||||
 DB 180 AGGTGACGTGACG 167

RESULT 11
 AAQ45597/C
 ID AAQ45597 standard; DNA; 423 BP.

AC AAQ45597;

DE 04-DEC-1993 (first entry)

DE Sequence encoding the VH of antibody B17X2.

XX Variable heavy antibody chain; human subgroup 4 germline; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 7..423

FT /+tag- a

XX WO9312231-A.

XX 24-JUN-1993.

XX 13-DEC-1991; 91WO-AU00583.

PR 13-DEC-1991; 91WO-AU00583.

XX (DOWC) DOW CHEM AUSTRALIA LTD.

PI Johnson KS, Mezes PS, Richard RA;

XX WPI: 1993-214173/26.

DR P-PSDB; AAR38315.

XX New composite antibody binding to tumour associated TAG-72

PT antigen - includes light chain variable region from human

PT subgroup 4 germline gene, useful, opt. as conjugate, for

PT diagnosis or treatment of cancer

XX Disclosure; Figure 4; 150pp; English.

CC Cell line B17X2 expresses an antibody utilising a variable light
 CC chain encoded by a gene derived from Hum4 VI, and a variable heavy
 CC chain which makes a stable VI and VH combination.

XX Sequence 423 BP; 98 A; 120 C; 102 G; 97 T; 6 other;

Query Match 56.0%; Score 14; DB 14; Length 423;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 aggtgcagctcagc 19
 |||||
 DB 133 AGGTGCAGCTCAGC 120

RESULT 12
 AA239427/C
 ID AA239427 standard; DNA: 423 BP.

AC AA239427;

DT 29-FEB-2000 (first entry)

XX Antibody B17X2 VH nucleotide sequence.

DE Tumor associated sialylated glycoprotein; TAG-72; cancer antigen;
 KM carcinoma lesion; diagnostic; cancer; antibody; human; B17X2;
 KM anti-mouse antibody hypersensitivity reaction; ss.

OS Homo sapiens.

XX US5976845-A.

PN 02-NOV-1999.

PD 07-JUN-1995; 950S-0487743.

PE 16-JUN-1994; 940S-0261354.

PR 19-APR-1990; 900S-0510697.

PT 20-OCT-1992; 920S-0964536.

XX (DOWC) DOW CHEM CO.

XX Johnson KS, Richard RA, Mezes PS;

DR WPI: 1999-619651/53.

XX P-PSDB; AAY57179.

PT Production of humanized anti-TAG-72 antibodies, used for the detection,
 in vivo imaging and treatment of cancers

XX Disclosure; Fig 4A-B; 85pp; English.

XX The invention relates to producing humanized anti-tumor associated
 CC sialylated glycoprotein (TAG-72) antibodies (anti-TAG). The antibodies
 CC have binding specificity for the cancer antigen TAG-72. These antibodies
 CC have variable regions with VL segments derived from human subgroup IV
 CC germ-line gene and a VH segment (encoded by the Vhalphatag germ-line gene)
 CC which is capable of combining with the VL to form a three dimensional
 CC structure having the ability to bind TAG-72. They can be used for the in
 CC vivo detection of carcinoma lesions. They can also be used for in vitro
 CC diagnostics. They can also be modified with therapeutic agents e.g. a
 CC radionuclide, drug, biological response modifier, toxin or another
 CC antibody for the treatment of cancers. The humanized anti-TAG-72
 CC antibodies can reduce harmful anti-mouse antibody hypersensitivity
 CC reactions.

SO Sequence 423 BP; 98 A; 120 C; 102 G; 97 T; 6 other;

Query Match 56.0%; Score 14; DB 20; Length 423;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 aggtgcagctcagc 19
 |||||
 DB 133 AGGTGCAGCTCAGC 120

RESULT 13
 AA23965/C
 ID AA23965 standard; DNA: 423 BP.
 XX

AC AA23965;
 XX 09-FEB-2000 (first entry)
 DT
 XX Human B17X2 antibody VH segment DNA.

DE Human; antibody; humanized; anti-tumor; sialylated glycoprotein antibody;
 KM TAG-72; cytostatic; cancer antigen; detection; carcinoma lesion;
 KM diagnostic; treatment; ss.

OS Homo sapiens.

XX US5976531-A.

PN 02-NOV-1999.

PD 16-JUN-1994; 940S-0261354.

PE 19-APR-1990; 900S-0510697.

PR 20-OCT-1992; 920S-0964536.

XX (DOWC) DOW CHEM CO.

XX Johnson KS, Mezes PS, Richard RA;

DR WPI: 1999-632731/54.

XX P-PSDB; AAY50688.

PT New humanized anti-TAG-72 antibodies, used for the detection, in vivo
 imaging and treatment of cancers

XX Disclosure; Figure 4A-B; 83pp; English.

XX This invention describes novel humanized anti-tumor associated
 CC sialylated glycoprotein antibodies (TAG-72) which have cytostatic
 CC activity. The antibodies have binding specificity for the cancer antigen
 CC TAG-72. They can be used for the in vivo detection of carcinoma lesions.
 CC They can also be used for in vitro diagnostics. They can also be modified
 CC with therapeutic agents e.g. a radionuclide, drug, biological response
 CC modifier, toxin or another antibody for the treatment of cancers. The
 CC humanized anti-TAG-72 antibodies can reduce harmful anti-mouse antibody
 CC hypersensitivity reactions.

SO Sequence 423 BP; 98 A; 120 C; 102 G; 97 T; 6 other;

Query Match 56.0%; Score 14; DB 20; Length 423;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 aggtgcagctcagc 19
 |||||
 DB 133 AGGTGCAGCTCAGC 120

RESULT 14
 AAT80181/C
 ID AAT80181 standard; DNA: 444 BP.
 XX
 AC AAT80181;

XX 02-APR-1998 (first entry)

DE Monoclonal antibody 105AD7 heavy chain variable region DNA sequence.

XX Monoclonal antibody 105AD7; complementarity determining region;

XX CDR; human; immune response; treatment; cancer; ss.

OS Homo sapiens.

XX Key Location/Qualifiers
 PH 1.444
 FT /*tag- a
 FT

FT /transl_except= (pos: 157..159, aa: Gly)
 FT /transl_except= (pos: 160..162, aa: Val)
 FT /Product= "105AD7 heavy chain variable region"
 PN MO9712021-A1.
 XX
 XX
 PD 04-SEP-1997.
 XX
 XX 28-FEB-1997; 97WO-GH00591.
 XX
 XX 29-FEB-1996; 96GB-0004321.
 PR 28-FEB-1996; 96GB-0004177.
 XX
 XX (CANC-) CANCER RES CAMPAIGN TECHNOLOGY LTD.
 PA
 XX Durrant IG, Robins RA, Spendlove I;
 PI
 DR WPI: 1997-448689/41.
 DR P-PSDB; AAMW6239.
 XX
 PT Nucleic acid encoding peptide derived from antibody 105AD7 - used to
 PT induce an immune response for treatment and prevention of cancer
 XX
 PS Claim 1; Fig 1A; 32pp; English.
 CC
 CC This is the DNA sequence of the monoclonal antibody 105AD7 heavy chain
 CC variable region. 105AD7 is a human monoclonal anti-idiotypic antibody
 CC that mimics T cell epitopes on the tumour associated antigen gp72 and
 CC has been used for therapeutic vaccination. A peptide derived from this
 CC antibody contains the 105AD7 heavy chain variable region and a 105AD7
 CC kappa chain. The complementarity determining regions (CDR) of this
 CC peptide and its fragments can stimulate immune responses. The peptide
 CC can be used for screening for its functional equivalents and mimetics.
 CC The peptides, and the fragments, equivalents and mimetics are used to
 CC stimulate an immune response to a tumour antigen, for the treatment or
 CC prevention of tumours. Transfected host cells are used to produce this
 CC peptide and the vectors can be used to generate the peptide in vivo for
 CC stimulation of an immune response. The peptide contains promiscuous
 CC helper epitopes which stimulate a response against tumour antigens other
 CC than gp72. These epitopes may help a cytotoxic T lymphocyte response to
 CC any co-injected antigen.
 CC
 SO Sequence 444 BP; 100 A; 134 C; 107 G; 103 T; 0 other;

Query Match 56.0%; Score 14; DB 18; Length 444;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 aggtgcaggtcagc 19
 |||
 DB 121 AGGTGCAGGTCAAGC 108

RESULT 15
 ID AA220406/c
 XX AA220406 standard; cDNA; 447 BP.
 AC
 XX AA220406;
 XX
 DT 19-NOV-1999 (first entry)
 XX
 DE IgG antibody 2.4.4 heavy chain coding sequence.
 XX
 KW Antibody; CD147; Igm; ABX-CBL; activated T-cell killing; leukaemia;
 KW activated B-cell; monocytic; graft versus host disease; therapy; cancer;
 KW organ transplant rejection disease; lymphoma; pancreatic disease;
 KW autoimmune disease; inflammatory disease; arthritis; binding site; ss.
 XX
 OS Homo sapiens.
 XX
 XX Key , Location/Qualifiers
 FT CDS 1..447

FT /*tag= a
 FT /note= "no stop codon given"
 FT
 PN WO9945031-A2.
 XX
 XX
 PD 10-SEP-1999.
 XX
 XX 03-MAR-1999; 99WO-US04583.
 XX
 XX 03-MAR-1998; 98US-0034607.
 PR 03-FEB-1999; 99US-0244253.
 XX
 XX (ABGE-) ABGENIX INC.
 PA
 XX Davis CG, Blacher RW, Corvatan JR, Culwell AR, Green LL, Hales J;
 PI Havrilla N, Ivanov VE, Lipani JA, Liu Q, Weber RF, Yang X;
 PI
 DR WPI: 1999-540816/45.
 DR P-PSDB; AAY34305.
 XX
 PT New monoclonal antibody, used for treating e.g. graft versus host
 PT disease, cancers, autoimmune diseases and inflammatory diseases
 XX
 PS Disclosure; Fig 30; 245pp; English.
 CC
 CC This sequence encodes the heavy chain of an antibody of the
 CC invention. The antibody is a monoclonal antibody (mAb) with an isotype
 CC that fixes complement and a variable region that binds to the epitope on
 CC CD147 bound by the Igm Mab ABX-CBL, providing that the antibody is not
 CC or resting or activated monocytes. The products and methods can be used
 CC for treating diseases involving activated T-cells or B-cells or
 CC monocytes, e.g. graft versus host disease (GVHD), organ transplant
 CC rejection diseases (e.g. renal transplant, ocular transplant), cancers
 CC (e.g. cancers of the blood (e.g. leukaemia's and lymphomas) and
 CC pancreatic), autoimmune diseases (e.g. lupus), and inflammatory diseases
 CC (e.g. arthritis).
 CC
 SO Sequence 447 BP; 96 A; 150 C; 106 G; 95 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 447;
 Best Local Similarity 100.0%; Pred. No. 93;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 aggtgcaggtcagc 19
 |||
 DB 34 AGGTGCAGGTCAAGC 21

RESULT 16
 ID AAQ78943/c
 XX AAQ78943 standard; DNA; 613 BP.
 AC
 XX AAQ78943;
 XX
 DT 07-AUG-1995 (first entry)
 XX
 DE Human immunoglobulin variable heavy chain #5.
 XX
 KW Primer; PCR; amplify; human; immunoglobulin; variable; heavy chain;
 KW cosmid; placenta; vector; pJB81; E.coli; mammalian; ds.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FT CDS 71..513
 FT /*tag= a
 FT /product= human immunoglobulin variable heavy chain
 FT intron 117..202
 FT /*tag= b
 FT misc-signal 514..516
 FT /*tag= c

RESULT 18
 AAAT5915/c
 ID AAAT5915 standard; cDNA; 1100 BP.
 AC AAAT5915;
 XX
 DT 12-JUN-2000 (first entry)
 DE Human protein clone HP10568 full length coding sequence.
 KW Human protein; hydrophobic domain; nutritional source; hematopoiesis;
 KW cytokine production; cell proliferation; cell differentiation;
 KW immune deficiency; infectious disease; autoimmune disorder; asthma;
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
 KW allergic reaction; osteoporosis; osteoarthritis; periodontal disease;
 KW Huntington's disease; liver fibrosis; Parkinson's disease;
 KW neurologic disease; tissue differentiation; contraceptive; stroke;
 KW coagulation disorder; myocardial infarction; inflammatory condition;
 KW septic shock; sepsis; ischemia; reperfusion injury; arthritis; tumour;
 KW nephritis; therapy; ss.
 OS Homo sapiens.
 XX
 XX MO200005367-A2.
 PN
 XX
 PD 03-FEB-2000.
 XX
 PF 22-JUL-1999; 99WO-JP03929.
 XX
 PR 24-JUL-1998; 98JP-0208820.
 PR 07-AUG-1998; 98JP-0224105.
 PR 25-AUG-1998; 98JP-0238116.
 PR 09-SEP-1998; 98JP-0254736.
 PR 29-SEP-1998; 98JP-0275505.
 XX
 PA (SAGA) SAGAMI CHEM RES CENT.
 PA (PROT-) PROTECENE INC.
 XX
 PI Kato S, Kimura T;
 DR WPI: 2000-182694/16.
 DR P-PSDB: AAY94857.
 XX
 PT Novel human proteins having hydrophobic domains useful for treating
 PT osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,
 PT multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -
 PS
 PS Claim 4; Page 203-205; 351pp; English.
 XX
 CC This sequence encodes a human protein of the invention, which has
 CC hydrophobic domains. The DNA sequences can be used as a probe or as a
 CC genetic marker. The protein can also be used as a marker, and to identify
 CC potential genetic disorders. The DNA and protein can also be used as
 CC nutritional sources or supplements. The protein exhibits cytokine, cell
 CC proliferation, cell differentiation activities and induces production of
 CC other cytokines in certain cell populations. The protein also exhibits
 CC immune stimulating or immune suppressing activity. It can be used in the
 CC treatment of various immune deficiencies and disorders, and to treat
 CC infectious diseases caused by viral, bacterial, fungal or other
 CC infections. The protein is also used for treating autoimmune disorders
 CC such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid
 CC arthritis. It is also useful in the treatment of allergic reactions and
 CC conditions such as asthma, and in immune suppression after organ
 CC transplantation. The protein is useful in regulation of haematopoiesis
 CC and consequently in the treatment of myeloid or lymphoid cell
 CC deficiencies. It is also used in compositions for tissue growth or
 CC regeneration. The protein is also used in the treatment of osteoporosis
 CC or osteoarthritis and in the treatment of periodontal disease and other
 CC tooth repair processes. The protein is used in the treatment of nervous
 CC system disorders such as Alzheimer's disease, Parkinson's disease, and
 CC Huntington's disease. They are useful for protection or regeneration and
 CC treatment of lung or liver fibrosis, reperfusion injury in various

CC tissues, and conditions resulting from systemic cytokine damage. They are
 CC also used for promoting or inhibiting tissue differentiation. They are
 CC also used as contraceptives since they exhibit activin or inhibin related
 CC activities and as a fertility inducing therapeutic. They are used for
 CC treating various coagulation disorders and in treatment and prevention of
 CC conditions resulting from coagulation activities e.g. myocardial
 CC infarction or stroke. They also acts as receptors, receptor ligands or
 CC inhibitors or agonists of receptor/ligand interactions. They are used to
 CC treat inflammatory conditions such as septic shock, sepsis, ischemia
 CC reperfusion injury, arthritis, and nephritis. They can be used to
 CC prevent tumours.
 XX
 XX Sequence 1100 BP; 221 A; 333 C; 314 G; 232 T; 0 other;
 Qy 6 agtgcaggtcagc 19
 Db 195 AGTGCAGGTCAAGC 182
 Query Match 56.0%; Score 14; DB 21; Length 1100;
 Best Local Similarity 100.0%; Pred. No. 89;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 19
 AAA39064/c
 ID AAA39064 standard; cDNA; 1130 BP.
 AC AAA39064;
 XX
 DT 30-AUG-2000 (first entry)
 DE Human secreted protein gene 13 SEQ ID NO:23.
 XX
 KW Human; secreted protein; cytosolic; anti-proliferative; vulnery;
 KW immunosuppressive; antibacterial; diagnosis; immune system; chemotaxis;
 KW hyperproliferative disorder; infectious disease; tissue regeneration;
 KW screening; food additive; preservative; wound healing;
 KW hyper-vascular disease; chromosome 11; ss.
 OS Homo sapiens.
 XX
 XX MO200017222-A1.
 PN
 XX
 PD 30-MAR-2000.
 XX
 PF 22-SEP-1999; 99WO-US22012.
 XX
 PR 23-SEP-1998; 98US-0101546.
 PR 02-OCT-1998; 98US-0102895.
 XX
 PA (HUMAN-) HUMAN GENOME SCI INC.
 PI Ruben SM, Rosen CA, Duan RD, Shi Y, Lafleur DW, Young PE, Ni J;
 PI Komatsoulis G, Endress GA, Soppet DR;
 DR WPI: 2000-283538/24.
 DR P-PSDB: AAB08903.
 XX
 PT Human secreted proteins and coding sequences useful in diagnostic and
 PT therapeutic methods for disorders such as immune system or
 PT proliferative disorders, related to the proteins -
 PS
 PS Claim 1; Page 329-330; 416pp; English.
 XX
 CC The polynucleotide sequences given in AAA39052 to AAA39088 encode the
 CC human secreted proteins given in AAB08891 to AAB08984. The human secreted
 CC proteins can have activities based on the tissues and cells they are
 CC expressed in. Examples of the activities are: cytosolic;
 CC anti-proliferative; immunosuppressive; antibacterial; and vulnery. The
 CC secreted proteins and their related polynucleotide sequences are useful
 CC for diagnostic and therapeutic methods useful for diagnosing and treating
 CC disorders related to the secreted proteins. The proteins, and

polynucleotide sequences may be useful for treating disorders of the immune system, hyperproliferative disorders, infectious disease, regeneration of tissues, for chemotaxis and for screening molecules that bind to the proteins. The proteins or polynucleotide sequences may be used as food additives or preservatives, to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, co-factors or other nutritional components. Agonists or antagonists of the proteins may be used to prevent scar tissue growth during wound healing, and hyper-vascular diseases. AAA39043 to AAA39051 and AAB08890 are sequences used in the exemplification of the present invention.

Sequence 1130 BP; 246 A; 334 C; 317 G; 233 T; 0 other;

Query Match
Best Local Similarity 100.0%; Pred. No. 89;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

6 aggtgcaggtcagc 19
|||||
201 AGGTGCAGGTCAAC 188

RESULT 20
AAA39087/c
ID AAA39087 standard; cDNA; 1135 BP.
XX
AC AAA39087;
XX
DT 30-AUG-2000 (first entry)
XX
DE Human secreted protein gene 13 SEQ ID NO:46.
XX
KW Human; secreted protein; cytosolic; anti-proliferative; vulnary;
KW immunosuppressive; antibacterial; diagnosis; immune system; chemotaxis;
KW hyperproliferative disorder; infectious disease; tissue regeneration;
KW screening; food additive; preservative; wound healing;
KW hyper-vascular disease; chromosome 11; ss.
XX
OS Homo sapiens.
XX
PN WO200017222-A1.
XX
PD 30-MAR-2000.
XX
PF 22-SEP-1999; 99WO-US22012.
XX
PR 23-SEP-1998; 98US-0101546.
PR 02-OCT-1998; 98US-0102895.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Rosen CA, Duan RD, Shi Y, Lafleur DW, Young PE, Ni J;
PI Komatsoulis G, Endress GA, Soppet DR;
XX
DR WPI; 2000-283538/24.
DR P-PSDB: AAB08926.
XX
PT Human secreted proteins and coding sequences useful in diagnostic and
PT therapeutic methods for disorders such as immune system or
PT proliferative disorders, related to the proteins
XX
PS Claim 1; Page 347-348; 416pp; English.
XX
CC The polynucleotide sequences given in AAA39052 to AAA39088 encode the
CC human secreted proteins given in AAB08891 to AAB08984. The human secreted
CC proteins can have activities based on the tissues and cells they are
CC expressed in. Examples of the activities are: cytosolic;
CC anti-proliferative; immunosuppressive; antibacterial; and vulnary. The
CC secreted proteins and their related polynucleotide sequences are useful
CC for diagnostic and therapeutic methods useful for diagnosing and treating
CC disorders related to the secreted proteins. The proteins, and

polynucleotide sequences may be useful for treating disorders of the immune system, hyperproliferative disorders, infectious disease, regeneration of tissues, for chemotaxis and for screening molecules that bind to the proteins. The proteins or polynucleotide sequences may be used as food additives or preservatives, to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, co-factors or other nutritional components. Agonists or antagonists of the proteins may be used to prevent scar tissue growth during wound healing, and hyper-vascular diseases. AAA39043 to AAA39051 and AAB08890 are sequences used in the exemplification of the present invention.

Sequence 1135 BP; 243 A; 333 C; 323 G; 234 T; 2 other;

Query Match
Best Local Similarity 100.0%; Pred. No. 89;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

6 aggtgcaggtcagc 19
|||||
188 AGGTGCAGGTCAAC 175

RESULT 21
AAA91019/c
ID AAA91019 standard; DNA; 1161 BP.
XX
AC AAA91019;
XX
DT 05-APR-2001 (first entry)
XX
DE Human secreted protein PRO7154 coding sequence.
XX
KW Secreted protein; human; PRO protein; neoplastic cell growth; tumour;
KW proliferation; leukemia; lymphoid malignancy; inflammatory disorder;
KW angiogenic disorder; immunologic disorder; PRO7154; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH CDS 102..1085
FT /*tag=a
FT /product= PRO7154
XX
PN WO200075317-A2.
XX
PD 14-DEC-2000.
XX
PF 15-MAY-2000; 2000MO-US13358.
XX
PR 09-JUN-1999; 99US-0138385.
PR 20-JUL-1999; 99US-0144790.
PR 03-AUG-1999; 99US-0146843.
PR 10-AUG-1999; 99US-0148188.
PR 17-AUG-1999; 99US-0149320.
PR 17-AUG-1999; 99US-0149327.
PR 17-AUG-1999; 99US-0149396.
PR 20-AUG-1999; 99US-0150114.
PR 31-AUG-1999; 99US-0151700.
PR 31-AUG-1999; 99US-0151734.
XX
PA (GETH) GENENTECH INC.
XX
PI Botstein DA, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
XX
DR WPI; 2001-071075/08.
DR P-PSDB: AAY97585.
XX
PT Antibodies against PRO polypeptides, useful for diagnosing and treating
PT tumours are associated with gene amplification, neoplastic cell growth
PT and proliferation in mammals -
XX

PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI: 2000-587534/55.
 XX P-PSDB: AAB53307.
 XX
 PT Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer -
 PS Claim 1: Page 520-521; 2104pp; English.
 XX
 CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB53307 to AAB54006. The
 CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
 CC vulnary, nephrotoxic, anti-infective and antibacterial activities, and
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins
 CC may also be used to prevent diseases such as neural disorders, immune
 CC system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, wounds, renal disorders, infectious
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 CC AAB54007 represent sequences used in the exemplification of the present
 CC invention.
 XX
 SO Sequence 1186 BP; 244 A; 361 C; 338 G; 242 T; 1 other;

Query Match 56.0%; Score 14; DB 21; Length 1186;
 Best Local Similarity 100.0%; Pred. No. 89;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 aggtgcaggtcagc 19
 |||
 DB 255 AGGTGCAGGTCAAC 242

RESULT 24
 AAT98726
 ID AAT98726 standard; DNA: 1191 BP.
 XX
 AC AAT98726;
 XX
 DT 09-NOV-1998 (first entry)
 XX
 DE DNA encoding a S. pneumoniae protein of unknown function.
 XX
 KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;
 KW immunological response; inoculation; antibody production; inhibitor;
 KW T cell immune response; antimicrobial compound; bacterial adhesion;
 KW extracellular matrix protein; protein-mediated cell invasion; wound;
 KW pathogenesis; ss.
 XX
 OS Streptococcus pneumoniae.
 XX
 FH Key Location/Qualifiers
 FT CDS 596..979
 FT /*tag= a
 FT
 XX
 XX MO9743303-A1.
 XX
 XX 20-NOV-1997.
 XX
 XX 14-MAY-1997; 97WO-US07950.
 XX
 XX 14-MAY-1996; 96US-0017670.
 XX
 PA (SMIK) SMITHKLINE BEECHAM CORP.

PA (SMIK) SMITHKLINE BEECHAM PLC.
 XX
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
 XX
 DR Stodola RK;
 XX
 DR WPI: 1998-008793/01.
 XX P-PSDB: AAW38682.
 XX
 PT Novel Streptococcus pneumoniae proteins and related DNA - useful for
 PT diagnosing anti-microbial agents for treatment of bacterial
 PT infections
 PS Claim 4: Pages 222-223; 483pp; English.
 XX
 CC This sequence encodes a Streptococcus pneumoniae protein of unknown
 CC function, and represents a DNA sequence of the invention.
 CC The DNA sequences were isolated from Streptococcus pneumoniae strain
 CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
 CC invention can be used to identify compounds which interact with and
 CC inhibit or activate the activity of the proteins. Antagonists can be
 CC used to treat diseases caused by S. pneumoniae proteins, through genetic
 CC immunisation. They can also be used to induce an immunological response
 CC in a mammal by inoculation with the S. pneumoniae proteins or delivery
 CC of the encoding nucleic acids in a vector adequate to produce antibody
 CC and/or T cell immune responses to protect the animal from disease. The
 CC proteins can also be used to identify antimicrobial compounds which are
 CC capable of inhibiting their bioactivity. In particular the proteins of
 CC the invention can be used to prevent adhesion of bacteria to mammalian
 CC extracellular matrix proteins on in-dwelling devices or in wounds, to
 CC block protein-mediated mammalian cell invasion, and to block the normal
 CC progression of pathogenesis in infections initiated other than by the
 CC implantation of in-dwelling devices or other surgical techniques.
 XX
 SO Sequence 1191 BP; 294 A; 258 C; 294 G; 345 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 1191;
 Best Local Similarity 100.0%; Pred. No. 89;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cgtcaggtcagc 15
 |||
 DB 954 cgtcaggtcagc 967

RESULT 25
 AAF44623/C
 ID AAF44623 standard; CDNA: 1368 BP.
 XX
 AC AAF44623;
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE Novel protein kinase cDNA, SEQ ID NO: 2.
 XX
 KW Human; mouse; protein kinase; antiarrhythmic; antisclerotic; osteopathic;
 KW immunosuppressive; cardiac; renal; anti-inflammatory; antiasthmatic;
 KW dermatological; antidiabetic; antifertility; gene therapy; vaccine;
 KW immune disorder; cardiovascular disease; neurodegenerative disease;
 KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;
 KW inflammatory pelvic disease; multiple sclerosis; psoriasis; ss.
 XX
 OS Mus musculus.
 XX
 XX MO200073469-A2.
 XX
 XX 07-DEC-2000.
 XX
 XX 26-MAY-2000; 2000MO-US14842.
 XX
 XX 28-MAY-1999; 99US-0136503.
 XX
 PA (SUGF-) SUGEN INC.

PS Claim 1; Page 60; 130pp; English.
 XX
 CC The sequence is that of a coding region isolated from
 CC S. pneumoniae. Its encoded protein, or agonists of it,
 CC may be useful as an antibacterial for treatment or
 CC prevention of infection, specifically caused by S. pneumoniae
 CC (particularly meningitis) but possibly also Helicobacter
 CC pylori (ulcers and gastric cancer). It may be of particular
 CC use before insertion of an in-dwelling device or any other
 CC invasive procedure. The protein, or nucleic acid encoding
 CC it, can also be used in vaccines to induce a cellular
 CC and/or humoral immune response, or to screen for other
 CC antibacterials. The DNA may also contain flanking sequences
 CC that are potential sources of control elements for bacterial
 CC gene expression. Detecting a sequence encoding the protein
 CC can be used diagnostically, e.g. to detect a mutation for
 CC serotyping or classifying infectious agents.
 XX
 SQ Sequence 1634 BP; 430 A; 339 C; 385 G; 480 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 1634;
 Best Local Similarity 100.0%; Pred. No. 87;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cgtcagtgacagtc 15
 |||||||||
 Db 954 cgtcagtgacagtc 967

RESULT 28

AAAI1992
 ID AAAI1992 standard; DNA; 37856 BP.

XX
 AC AAAI1992;

XX
 DT 07-ANG-2000 (first entry)

XX S. cellulorum DNA encoding polyketide and heteropolyketide enzymes.
 XX
 XX Polyketide biosynthesis; heteropolyketide biosynthesis; mutasynthetases;
 KM epothilone; cytotoxic; immunosuppressant; antibiotic; antifungal;
 KM plant-protection; ds.
 XX
 OS Sorangium cellulosum.

XX
 FH Key Location/Qualifiers
 FT complement (3398..6100)

FT /*tag= a

FT /product= "ORF1-tRNA synthetase"

FT /note= "gtg start codon"

FT complement (6374..7111)

FT /*tag= b

FT /product= "ORF2-monoxygenase"

FT complement (8433..9550)

FT /*tag= c

FT /product= "ORF3-aminotransferase"

FT /note= "AGT start codon given in the specification"

FT 9855..11393

FT /*tag= d

FT /product= "ORF4-tyrosine/DOPA-Decarboxylase"

FT /note= "GTG start codon"

FT 12212..13658

FT /*tag= e

FT /product= "ORF5-3-oxoacyl-ACP-reductase"

FT /note= "ACC start codon"

FT 15374..19984

FT /*tag= f

FT /product= "ORF6-polyketide synthase"

FT 20003..27889

FT /*tag= g

FT /product= "ORF7-peptide synthetase"

FT 28251..29400

FT CDS

FT /*tag= h
 FT /product= "ORF8-transpeptidase"
 FT complement (30040..31720)
 FT /*tag= i
 FT /product= "ORF9-regulation element"
 FT /note= "CGC stop codon"
 FT 31982..32932
 FT /*tag= j
 FT /product= "ORF10-transcription regulator"
 FT 33128..33613
 FT /*tag= k
 FT /product= "ORF11-regulation element"
 FT /note= "GTG start codon"
 FT 33661..34077
 FT /*tag= l
 FT /product= "ORF12-regulation element"
 FT complement (35255..35616)
 FT /*tag= m
 FT /product= "ORF13-transcription regulator"
 FT complement (35730..36242)
 FT /*tag= n
 FT /product= "ORF14-transcription regulator"
 FT /note= "GTG start codon"

DE19846493-A1.

13-APR-2000.

XX 09-OCT-1998; 98DE-1046493.

XX 09-OCT-1998; 98DE-1046493.

XX (GBPB) GBS BIOTECHNOLOGISCHE FORSCHUNG MBH.

XX Beyer S, Mueller R;

XX WPI: 2000-294101/26.

XX DNA sequence coding for products involved in the biosynthesis of
 PT polyketide or heteropolyketide compounds, especially epothilone

XX Claim 3; Page 20-33; 36pp; German.

XX This invention describes a novel DNA sequence (I) whose expression
 CC products effect or are involved in the enzymatic biosynthesis,
 CC mutasynthesis or partial synthesis of polyketide or heteropolyketide
 CC compounds (II). (I) can be inserted into an expression vector and used
 CC to transform or transfect prokaryotic or eukaryotic cells with the aim
 CC of obtaining strains that produce large amounts of polyketide or
 CC heteropolyketide compounds, especially epothilones, which have cytotoxic
 CC and/or immunosuppressant and antibiotic and antifungal activities and
 CC are useful as plant-protection agents. This sequence represents the DNA
 CC sequence isolated from Sorangium cellulosum which is described in the
 CC method of the invention.

XX Sequence 37856 BP; 5655 A; 13666 C; 12913 G; 5622 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 37856;
 Best Local Similarity 100.0%; Pred. No. 76;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 gtcagtgacagtc 16
 |||||||||
 Db 35785 gtcagtgacagtc 35798

RESULT 29

AAV30458

ID AAV30458 standard; DNA; 534720 BP.

XX
 AC AAV30458;

XX

```

DT 14-OCT-1998 (first entry)
XX
XX Rhizobium species plasmid pNGR234a.
XX
XX Symyiosis: open reading frame; ORF: plasmid; vector: transportation;
KW degradation; metabolism; host range; nitrogen fixation; nodulation;
KW legume; plant; ds.
XX
XX Rhizobium sp.
XX
XX Key Location/Qualifiers
FH 417796..418671
CDS
FT /tag= a
FT /standard_name= "ORF K1"
FT /product= "oligopeptide permease"
FT /note= "homologous to the Oppc gene"
FT 418673..419680
CDS
FT /tag= b
FT /standard_name= "ORF K2"
FT /product= "oligopeptide permease"
FT /note= "homologous to the Oppd gene"
FT 419677..420738
CDS
FT /tag= c
FT /standard_name= "ORF K3"
FT /product= "oligopeptide permease"
FT /note= "homologous to the Oppf gene"
FT 420774..422159
CDS
FT /tag= d
FT /standard_name= "ORF K4"
FT /product= "encapsulation-like protein"
FT /note= "homologous to the Capa gene"
FT 422628..424031
CDS
FT /tag= e
FT /standard_name= "ORF K5"
FT /product= "aminotransferase-like protein"
FT /note= "homologous to the Bida gene"
FT 424056..425594
CDS
FT /tag= f
FT /standard_name= "ORF K6"
FT /product= "(semi)aldehyde dehydrogenase-like protein"
FT /note= "complement (426949..428028)"
FT 428292..429623
CDS
FT /tag= g
FT /standard_name= "ORF K7"
FT /product= "transposase homologue"
FT /note= "homologous to the Tnp gene"
FT 429623
CDS
FT /tag= h
FT /standard_name= "ORF K8"
FT /product= "glutamate dehydrogenase-like protein"
FT /note= "homologous to the GUDI gene"
FT 430538..431284
CDS
FT /tag= i
FT /standard_name= "ORF K9"
FT /product= "transposase homologue"
FT /note= "complement (431296..432840)"
FT 432840
CDS
FT /tag= j
FT /standard_name= "ORF K10"
FT /product= "transposase homologue"
FT /note= "homologous to the Tnp gene"
FT 433880..434110
CDS
FT /tag= k
FT /standard_name= "ORF K11"
FT /product= "protein of unknown function"
FT /note= "homologous to the Fixu gene"
FT 434107..434433
CDS
FT /tag= l
FT /standard_name= "ORF K12"
FT /product= "protein of unknown function"
FT /note= "complement (434517..434711)"
FT 434711
CDS
FT /tag= m
FT /standard_name= "ORF K13"
FT /product= "ferrodoxin/ferrodoxin-like protein"
FT /note= "homologous to the fdxn gene"
FT
FT complement (434753..436234)
FT /tag= n
FT /standard_name= "ORF K14"
FT /gene= "nifb"
FT /product= "protein involved in FeMo co-factor
FT biosynthesis"
FT complement (436460..438130)
CDS
FT /tag= o
FT /standard_name= "ORF K15"
FT /gene= "nifA"
FT /product= "positive regulator of nif, fix and other
FT genes"
FT complement (438297..438590)
CDS
FT /tag= p
FT /standard_name= "ORF K16"
FT /gene= "fixX"
FT /product= "protein required for nitrogenase activity"
FT complement (438605..439912)
CDS
FT /tag= q
FT /standard_name= "ORF K17"
FT /gene= "fixC"
FT /product= "protein required for nitrogenase activity"
FT complement (439923..441032)
CDS
FT /tag= r
FT /standard_name= "ORF K18"
FT /gene= "fixB"
FT /product= "protein required for nitrogenase activity"
FT complement (441042..441899)
CDS
FT /tag= s
FT /standard_name= "ORF K19"
FT /gene= "fixA"
FT /product= "protein required for nitrogenase activity"
FT complement (442316..442636)
CDS
FT /tag= t
FT /standard_name= "ORF K20"
FT /product= "protein of unknown function"
FT complement (443313..443879)
CDS
FT /tag= u
FT /standard_name= "ORF K21"
FT /product= "protein of unknown function"
FT 444337..445029
CDS
FT /tag= v
FT /standard_name= "ORF K22"
FT /product= "ferrodoxin-like protein"
FT /note= "homologous to the Nifv gene"
FT 445088..446602
CDS
FT /tag= w
FT /standard_name= "ORF K23"
FT /gene= "dctA"
FT /product= "C4-dicarboxylate transport protein"
FT /note= "homologous to the DctAI gene"
FT 446599..447843
CDS
FT /tag= x
FT /standard_name= "ORF L1"
FT /product= "cytochrome P450-like protein"
FT /note= "homologous to the Camc gene"
FT 447844..448500
CDS
FT /tag= y
FT /standard_name= "ORF L2"
FT /product= "gamma-hexachlorocyclohexane-dechlorinase-like
FT protein"
FT /note= "homologous to the LinA gene"
FT 448497..450203
CDS
FT /tag= z
FT /standard_name= "ORF L3"
FT /product= "putative protein with degradative function"
FT 450341..451396
CDS
FT /tag= aa
FT /standard_name= "ORF L4"
FT /product= "luciferase alpha-subunit-like protein"
FT /note= "homologous to the luxA gene"
FT 452960..453494
FT /tag= ab

```

```

FT /standard_name= "ORF L6"
FT /gene= "nifD"
FT /product= "alpha-subunit of Fmo protein of nitrogenase"
FT CDS 454590..456131
FT /tag= ac
FT /standard_name= "ORF L7"
FT /gene= "nifK"
FT /product= "beta-subunit of Fmo protein of nitrogenase"
FT CDS 456187..457677
FT /tag= ad
FT /standard_name= "ORF L8"
FT /product= "protein involved in Fmo co-factor
FT biosynthesis"
FT /note= "homologous to the NifB gene"
FT CDS 457687..459096
FT /tag= ae
FT /standard_name= "ORF L9"
FT /product= "protein involved in Fmo co-factor
FT biosynthesis"
FT /note= "homologous to the FixF gene"
FT CDS 459093..459575
FT /tag= af
FT /standard_name= "ORF L10"
FT /product= "protein of unknown function"
FT /note= "homologous to the NifX gene"
FT CDS 459579..460067
FT /tag= ag
FT /standard_name= "ORF L11"
FT /product= "protein of unknown function"
FT /note= "homologous to the NifX gene"
FT CDS 460501..460920
FT /tag= ah
FT /standard_name= "ORF L12"
FT /product= "protein similar to part of the Fe protein
FT of nitrogenase"
FT /note= "homologous to the NifH gene"
FT CDS 461228..461545
FT /tag= ai
FT /standard_name= "ORF L13"
FT /product= "protein of unknown function"
FT CDS 463201..464739
FT /tag= aj
FT /standard_name= "ORF L14"
FT /product= "peptidase-like protein"
FT /note= "homologous to the Df-MFP gene"
FT CDS 464736..466079
FT /tag= ak
FT /standard_name= "ORF L15"
FT /product= "processing protease-like protein"
FT /note= "homologous to the PP gene"
FT CDS 466590..467021

Query Match 56.0%; Score 14; DB 19; Length 534720;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tcaggtgcaggtca 17
|||||
Db 435493 tcaggtgcaggtca 435506

RESULT 30
AAV30459
ID AAV30459 standard; DNA: 536165 BP.
XX
AC AAV30459;
XX
XX 06-JUL-1999 (first entry)
DE Rhizobium species symbiotic plasmid pNGR234.
XX
XX Symbiosis; open reading frame; ORF; plasmid; vector; transportation;
KW degradation; metabolism; host range; nitrogen fixation; nodulation;
KW legume; plant; ds.

```

```

XX XX Rhizobium sp.
OS W09802560-A2.
PN 22-JAN-1998.
XX
PD 10-JUL-1997; 97MO-IB00950.
XX
PF 20-MAY-1997; 97GB-0010395.
PR 12-JUL-1996; 96EP-0730001.
XX
PA (MOLE-) INST MOLECULAR BIOTECHNOLOGY.
XX (BIOL-) LAB BIOLOGIE MOLECULAIRE PLANTES SUPERIE.
PI Broughton WJ, Freiberg CB, Perret XP, Rosenthal A;
XX WPI: 1998-110606/10.
XX
XX New isolated symbiotic plasmid from Rhizobium sp. NGR234 - used to
PT develop products for modifying plant characteristics, e.g. nitrogen
PT fixation, synthesis of compounds and stress response
XX
XX Claim 1; Fig 3; 228bp; English.
XX
XX This is the nucleotide sequence of the plasmid pNGR234a isolated from
CC Rhizobium sp. NGR234. Open reading frames (ORF) derivable from the
CC nucleotide sequence are claimed. The nucleotide sequences or ORFs can
CC be used e.g. in the transportation of compounds to and from an organism
CC which is a host to at least one of the nucleotide sequences, ORFs or
CC proteins, the degradation and/or metabolism of organic, inorganic,
CC natural or xenobiotic substances in a host organism or the modification
CC of the host range, nitrogen fixation abilities; for obtaining a synthetic
CC minimal set of ORFs required for functional Rhizobium-legume symbiosis,
CC especially for nodulation efficiency on host plants.
XX
SQ Sequence 536165 BP; 111291 A; 155755 C; 157864 G; 111255 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 536165;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tcaggtgcaggtca 17
|||||
Db 435493 tcaggtgcaggtca 435506

RESULT 31
AAO13647
ID AAO13647 standard; DNA: 18 BP.
XX
AC AAO13647;
XX
XX 28-NOV-1991 (first entry)
DE Exon B/intron junction of PTPase B1 gene.
XX
XX PTPB1; protein phosphotyrosyl phosphatase B1;
KW growth suppression activity; ANTL; myeloproliferation; ss.
XX
XX Homo sapiens.
XX
XX key Location/Qualifiers
FH key 1..9
FT exon /*tag= a
FT intron /note= "last 3 codons of exon B"
FT 10..18
FT /*tag= b
FT /note= "5' end of intron"
XX
XX W09113173-A.
XX

```

PD 05-SEP-1991.
XX
XX 01-MAR-1991; 91WO-0001432.
XX
XX 18-JAN-1991; 91US-0643041.
PR 02-MAR-1990; 90US-0487733.
XX
XX (BIOF-) APPL BIOTECHN INC.
PA
XX
XX Bruskun AM, Hill DE;
XX
XX WPI; 1991-281485/38.
DR
XX
XX Vector confg. protein phospho-tyrosyl phosphatase gene fragment -
PT for use in diagnosis of cancers, e.g. acute non-lymphocytic
leukemia
PS
XX
XX Disclosure; Fig 4B; 73pp; English.
XX
XX The PTase B1 genomic sequence comprises a number of exons and
CC introns. Exon B encodes amino acids Met(235) to Gln(288) of the
CC PTase B1 protein (see AKR1907). This is the junction of the 3' end
CC of exon B and the 5' end of the intron.
CC See AAQ13643-Q13654.
CC
XX
SQ Sequence 18 BP; 3 A; 5 C; 5 G; 5 T; 0 other;

Query Match 52.0%; Score 13; DB 12; Length 18;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 gtgcaggtcagca 20
Db 4 gtgcaggtcagca 16
|||||

RESULT 32
AAK85925/C
ID AAK85925 standard; DNA; 123 BP.
XX
XX AAK85925;
AC
XX
XX 13-SEP-1999 (first entry)
DT
XX
XX Oligonucleotide used to produce heavy chain variable region of Ig NEW.
DE
XX
XX Light chain variable region; interleukin-4; IL-4; antibody 3B9;
KW chimeric antibody; humanised antibody; IL-4 mediated allergic reaction;
KW immunoglobulin E-mediated allergic reaction; allergic rhinitis;
KW conjunctivitis; atopic dermatitis; atopic asthma; anaphylactic shock;
KW rheumatoid arthritis; host-versus-graft disease; renal disease;
KW allergy; Ig NEW; ss.
XX
XX Synthetic.
OS
XX
XX US5928904-A.
PM
XX
XX 27-JUL-1999.
PD
XX
XX 07-JUN-1995; 95US-0483632.
PF
XX
XX 07-JUN-1995; 95US-0483632.
PR 07-SEP-1993; 93US-0117366.
PR 14-OCT-1993; 93US-0136783.
PR 07-SEP-1994; 94WO-US10308.
XX
XX (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
XX
XX Gross MS, Holmes SD, Sylvester DR;
PI
XX
XX WPI; 1999-429500/36.

XX
XX New DNA molecules encoding recombinant antibodies useful for
PT treating IL4-mediated conditions
PT
XX
XX Disclosure; Columns 61-63; 50pp; English.
PS
XX
XX The specification describes chimeric and humanised IL-4
CC monoclonal antibodies. The antibodies of the invention are used in
CC therapeutic and pharmaceutical compositions for treating IL-4 mediated
CC and immunoglobulin E-mediated allergic reactions e.g. allergic rhinitis,
CC conjunctivitis, atopic dermatitis, atopic asthma, anaphylactic shock,
CC rheumatoid arthritis, host-versus-graft disease and renal disease.
CC They are also useful in the diagnosis of an allergy or condition
CC associated with excess IL-4 production through the measurement e.g. by
CC ELISA of circulating endogenous IL-4 levels in humans. Oligonucleotides
CC AAK85925-28 were used to produce the heavy chain variable region of
CC Ig NEW, in the course of the invention.
CC
XX
SQ Sequence 123 BP; 25 A; 37 C; 41 G; 20 T; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 123;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 agtgcaggtcag 18
Db 70 AGGTGACAGTCCAG 58
|||||

RESULT 33
AAQ60177/C
ID AAQ60177 standard; DNA; 225 BP.
XX
XX AAQ60177;
AC
XX
XX 16-MAR-1994 (first entry)
DT
XX
XX Human brain Expressed Sequence Tag EST02167.
DE
XX
XX Gene transcription product; genetic markers; tagging; in vivo;
KW transcription; mapping; locations; chromosomes; chromosomal; ss.
XX
XX Homo sapiens.
OS
XX
XX WO9316178-A.
PN
XX
XX 19-AUG-1993.
PD
XX
XX 12-FEB-1993; 93WO-US01294.
PR
XX
XX 12-FEB-1992; 92US-0837195.
PR
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICE.
PA
XX
XX Adams MD, Moreno RF, Venter CJ;
PI
XX
XX WPI; 1993-272882/34.
DR
XX
XX Enriched oligonucleotides and corresp. sequences - used as
PT markers for human genes transcribed in-vivo, facilitate tagging
PT of most human genes
XX
XX Example 4; Page 293; 500pp; English.
PS
XX
XX The Expressed Sequence Tag was isolated from a human brain cDNA
CC library as part of a large set of ESTs which can be used as markers
CC for human genes transcribed in vivo. They can be used to facilitate
CC tagging of most human genes, for mapping locations of expressed genes
CC on chromosomes, for individual or forensic identification, for mapping
CC locations of disease-associated genes, for identification of tissue
CC type, and for prepn. of antisense sequences, probes and constructs.
CC EST02167 has a "poor" coding probability as evaluated using the

CC coding-region prediction program CRM. See also AAG59041-061440.
 XX Sequence 225 BP; 67 A; 67 C; 49 G; 41 T; 1 other;

Query Match 52.0%; Score 13; DB 14; Length 225;
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 tgcagtcagcagc 21
 |||||
 DB 96 TGCAGTCAGCAGC 84

RESULT 34
 AAC00049/c
 ID AAC00049 standard; cDNA; 276 BP.

XX AAC00049;
 AC AAC00049;
 XX 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 47.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KM gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EPI033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Mline Edwards J, Duclert A, Giordano J;

XX WPI: 2000-500381/45.

XX P-PSDB: AAG00043.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

PS Claim 1; SEQ ID 47; 71pp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.

SO Sequence 276 BP; 55 A; 93 C; 69 G; 59 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 276;

Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ggltgcagtcagc 19

DB 143 GGTGCGAGTCAGC 131
 |||||

RESULT 35
 AAX85929/c
 ID AAX85929 standard; DNA; 363 BP.

XX AAX85929;

XX 13-SEP-1999 (first entry)

DE DNA encoding the heavy chain variable region of Ig NEM.

KW Heavy chain variable region; interleukin-4; IL-4; antibody 3B9;
 KM chimeric antibody; humanised antibody; IL-4 mediated allergic reaction;
 KM immunoglobulin E-mediated allergic reaction; allergic rhinitis;
 KM conjunctivitis; atopic dermatitis; atopic asthma; anaphylactic shock;
 KM rheumatoid arthritis; host-versus-graft disease; renal disease;
 KM allergy; Ig NEM; ss.

XX Synthetic.

XX US5928904-A.

XX 27-JUL-1999.

XX 07-JUN-1995; 95US-0483632.

XX 07-JUN-1995; 95US-0483632.

XX 07-SEP-1993; 93US-0117366.

XX 14-OCT-1993; 93US-0136783.

XX 07-SEP-1994; 94WO-US10308.

XX (SMK) SMITHKLINE BECHAM CORP.

XX (SMK) SMITHKLINE BECHAM PLC.

XX Gross MS, Holmes SD, Sylvester DR;

XX WPI: 1999-429500/36.

XX P-PSDB: AAY23780.

PT New DNA molecules encoding recombinant antibodies useful for
 PT treating IL4-mediated conditions

PS Disclosure; Columns 63-65; 50pp; English.

CC The specification describes chimeric and humanised IL-4
 CC monoclonal antibodies. The antibodies of the invention are used in
 CC therapeutic and pharmaceutical compositions for treating IL-4 mediated
 CC and immunoglobulin E-mediated allergic reactions e.g. allergic rhinitis,
 CC conjunctivitis, atopic dermatitis, atopic asthma, anaphylactic shock,
 CC rheumatoid arthritis, host-versus-graft disease and renal disease.
 CC They are also useful in the diagnosis of an allergy or condition
 CC associated with excess IL-4 production through the measurement e.g. by
 CC ELISA of circulating endogenous IL-4 levels in humans. The present
 CC sequence encodes the heavy chain variable region of Ig NEM, and is
 CC used in the course of the invention.

XX Sequence 363 BP; 79 A; 116 C; 111 G; 57 T; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 363;

Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtgcagtcagc 18
 |||||

DB 70 AGGTGCGAGTCAGC 58

RESULT 36

AAX79527/c

ID AAX79527 standard; cDNA; 363 BP.
 XX
 AC AAX79527;
 XX
 DT 11-AUG-1999 (first entry)
 XX
 DE Heavy chain coding sequence for humanised 3B9 antibody.
 XX
 KM Antibody; interleukin-4, IL4; immunoglobulin E; IgE mediated disease;
 KM allergic disorder; allergic rhinitis; conjunctivitis; atopic dermatitis;
 KM atopic asthma; anaphylactic shock; cell proliferation regulator; therapy;
 KM autoimmune disease; graft versus host disease; ss.
 XX
 OS Synthetic.
 XX
 PN US5914110-A.
 PD 22-JUN-1999.
 XX
 PF 07-JUN-1995; 95US-0483636.
 XX
 PR 07-JUN-1995; 95US-0483636.
 PR 07-SEP-1993; 93US-0117366.
 PR 14-OCT-1993; 93US-0136783.
 PR 07-SEP-1994; 94WO-US10308.
 XX
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 XX
 PI Gross MS, Holmes SD, Sylvester DR;
 DR WP1; 1999-370482/31.
 DR P-PSDB; AAY18122.
 XX
 PT Recombinant IL4 antibodies
 PS
 PS Example 3; Column 63-64; 50pp; English.
 XX
 CC This sequence encodes the heavy chain of the humanised 3B9
 CC antibody of the invention. The antibody is a chimeric or
 CC humanised interleukin-4 (IL4) monoclonal antibody for the treatment of
 CC immunoglobulin E (IgE) mediated diseases. The antibodies are useful for
 CC the treatment of allergic disorders such as allergic rhinitis,
 CC conjunctivitis, atopic dermatitis, atopic asthma and anaphylactic shock.
 CC The antibodies are also useful for regulating B and T cell proliferation
 CC and as such are useful in the treatment of autoimmune diseases and graft
 CC versus host disease.
 XX
 SO Sequence 363 BP; 79 A; 116 C; 111 G; 57 T; 0 other;
 XX

Query Match 52.0%; Score 13; DB 20; Length 363;
 Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 aggtgcaggtcag 18
 |||||||||
 DB 70 AGGTGCAGGTGAC 58

RESULT 37
 AAF22004
 ID AAF22004 standard; DNA; 375 BP.
 XX
 AC AAF22004;
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE Human breast and ovarian cancer associated antigen gene SEQ ID 391.
 XX
 KM Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 KM neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KM antidiabetic; antiinflammatory; antitumor; vulnerary; anticonvulsant;
 KM

KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KW cardiovascular disorder; wound healing; neurological disease; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200055173-A1.
 PD 21-SEP-2000.
 XX
 PR 08-MAR-2000; 2000WO-US05881.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 DR WP1; 2000-611515/58.
 DR P-PSDB; AAB59101.
 XX
 PT New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases -
 XX
 PS Claim 1; Page 823; 1299pp; English.
 XX
 CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;
 CC neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 CC antidiabetic; antiinflammatory; antitumor; vulnerary; anticonvulsant;
 CC antibacterial; antitumor; antiparasitic and cardiant activity. The
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,
 CC particularly breast and ovarian cancer. The nucleic acid sequences,
 CC proteins, agonists and antagonists may also be used in the diagnosis,
 CC prevention and treatment of immune disorders e.g. Addison's disease,
 CC allergies, autoimmune hemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, Rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemia; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases.
 XX
 SO Sequence 375 BP; 79 A; 77 C; 129 G; 71 T; 19 other;
 XX

Query Match 52.0%; Score 13; DB 21; Length 375;
 Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 gtcaagtcaggt 15
 |||||||||
 DB 171 gtcaagtcaggt 183

RESULT 38
 AAT61237/c
 ID AAT61237 standard; DNA; 378 BP.
 XX
 AC AAT61237;
 XX
 DT 13-MAY-1997 (first entry)
 XX
 DE Human anti-RSV monoclonal antibody RF-2 VH domain DNA.
 XX
 KM Monoclonal antibody; MAb; RF-1; RF-2; respiratory syncytial virus;
 KM

XX Dumas Milne Edwards J, Duclert A, Giordano J;
 XX WPI: 2000-500381/45.
 DR P-PSDB; AAG00027.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 1: SEQ ID 31; 71pp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA⁺ RNAs
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.

XX Sequence 400 BP; 90 A; 117 C; 89 G; 92 T; 12 other;

Query Match 52.0%; Score 13; DB 21; Length 400;
 Best Local Similarity 100.0%; Pred. No. 3e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtgcaggtcag 18
 |||||
 Db 146 AGGTGCAGGTCTCAG 134

RESULT 41

AAK79207/c

ID AAK79207 standard; cDNA: 423 BP.

XX AAK79207;

XX 17-AUG-1999 (first entry)

XX Anti-HIV-1 gp120 antibody 694-D VH chain gene.

XX Heavy chain; variable region; human; HIV-1; gp120; monoclonal antibody;

XX epitope; V3 loop; heterohybridoma; human Immunodeficiency virus-1;

XX peripheral blood lymphocyte; Epstein-Barr virus; EBV; AIDS; ds.

XX Homo sapiens.

XX US5914109-A.

XX 22-JUN-1999.

XX 21-NOV-1994; 94US-0345321.

XX 23-APR-1992; 92US-0872675.

XX 15-JUN-1990; 90US-0538451.

XX 12-APR-1991; 91US-0684090.

XX 21-NOV-1994; 94US-0345321.

XX (UYNX) UNIV NEW YORK STATE.

XX Gorny MK, Zolla-Pazner S;

XX WPI: 1999-370481/31.

XX P-PSDB; AAY07486.

PT Heterohybridoma producing human monoclonal antibodies to human
 PT immunodeficiency virus-1
 XX
 XX Example 8; Fig 13; 42pp; English.

CC This sequence represents the coding region for the heavy chain variable
 CC region of the human anti-HIV-1 gp120 monoclonal antibody 694-D. The
 CC antibody is targeted to an epitope on the V3 loop of gp120. The
 CC invention relates to the generation of heterohybridomas producing human
 CC monoclonal antibodies to a neutralising epitope of human immunodeficiency
 CC virus-1 (HIV-1) prepared by transfecting peripheral blood lymphocytes
 CC with Epstein-Barr virus. The antibodies can be used to treat someone
 CC infected with HIV-1 or suffering from AIDS.

XX Sequence 423 BP; 96 A; 127 C; 104 G; 95 T; 1 other;

Query Match 52.0%; Score 13; DB 20; Length 423;
 Best Local Similarity 100.0%; Pred. No. 3e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtgcaggtcag 18
 |||||
 Db 127 AGGTGCAGGTCTCAG 115

RESULT 42

AAZ42289/c

ID AAZ42289 standard; cDNA: 437 BP.

XX AAZ42289;

XX 01-FEB-2000 (first entry)

XX Human 5' EST isolated from a cDNA library SEQ ID NO:48.

XX Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;

XX gene therapy; chromosome mapping; upstream regulatory sequence;

XX forensic; location; development; protein synthesis; stability;

XX regulation; identification; ss.

XX Homo sapiens.

XX WO9953051-A2.

XX 21-OCT-1999.

XX 09-APR-1999; 99WO-1B00712.

XX 09-APR-1998; 98US-0057719.

XX 28-APR-1998; 98US-0069047.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI: 2000-038446/03.

XX P-PSDB; AAY64675.

XX Novel secreted protein 5' expressed sequence tag sequences used in

XX diagnostic, forensic, gene therapy, and chromosome mapping procedures

XX Claim 1; Page 198; 837pp; English.

XX AAZ42265 to AAZ43075 represent novel 5' expressed sequence tag (EST)

XX sequences, corresponding to human secreted proteins. AAY64651 to

XX AAY63438 represent the EST-related proteins corresponding to AAZ42265 to

XX AAZ43052. The 5' ESTs can be used for producing secreted human gene

XX products. They can be used to identify and isolate 5' untranslated

XX regions (UTRs) and upstream regulatory regions which control the

XX location, development stage, rate, and quantity of protein synthesis, as

XX well as stability of mRNA. The ESTs are also useful as probes for

XX chromosome mapping, and to obtain full length cDNA clones. The ESTs can

CC also be used in forensic procedures to identify individuals, or in
 CC diagnostic procedures to identify individuals having genetic diseases
 CC resulting from abnormal gene expression. The products may also be used in
 CC gene therapy protocols. The nucleic acids encoding signal peptides can be
 CC used for directing extracellular secretion of a polypeptide or the
 CC insertion of a polypeptide into a membrane, or importing a polypeptide
 CC into a cell. The proteins encoded by the EST sequences may be useful in
 CC treating a variety of human conditions. Secreted proteins have
 CC therapeutic value, and the identification of new secreted proteins is
 CC valuable. AA242249 to AA242264 and AA242264 to AA242265 represent
 CC sequences used in the exemplification of the present invention.

XX
 SQ Sequence 437 BP; 94 A; 132 C; 105 G; 102 T; 4 other;

Query Match 52.0%; Score 13; DB 21; Length 437;
 Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 aggtgcaggtcag 18
 |||||
 DB 146 AGGTGCAGGTGAC 134

RESULT 43
 AACT5419
 ID AACT5419 standard; cDNA; 472 BP.
 XX
 AC AACT5419;
 XX
 DT 08-FEB-2001 (first entry)
 DE Human OREF974 polynucleotide sequence SEQ ID NO:1947.

XX Human; open reading frame; OREF; detection: cytostatic; hepatotropic;
 KM vulnery; antiprositic; antiparkinsonian; neurotropic; neuroprotective;
 KM anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
 KM immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KM hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KM antiviral; antibacterial; antifungal; antineumatic; antihydroid;
 KM antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KM neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KM cholesterol ester storage; systemic lupus erythematosus; infection;
 KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KM bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KM thrombosis; contraceptive; ss.

XX Homo sapiens.
 OS
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCM-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Leach M;
 XX
 DR WPI: 2000-602362/57.
 DR P-PSDB; AABA1210.

XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX

PS Claim 5; Page 1479-1480; 5507pp; English.

XX AACT74446 to AACT7606 encode the proteins given in ABA40237 to ABA43397,
 CC which represent the human OREF open reading frames 1 to 3161. The OREF
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antiprositic; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;
 CC antihydroid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an OREF-associated disorder. The
 CC nucleic acids can be used to express OREF proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 472 BP; 94 A; 151 C; 174 G; 53 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 472;
 Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 caggtgcaggtca 17
 |||||
 DB 185 caggtgcaggtca 197

RESULT 44
 AA220408/c
 ID AA220408 standard; cDNA; 477 BP.
 XX
 AC AA220408;
 XX
 DT 19-NOV-1999 (first entry)
 DE Igg antibody 2.3.2 heavy chain coding sequence.

XX Antibody; CD147; Igm; ABX-CBL; activated T-cell killing; leukaemia;
 KM activated B-cell; monocyte; graft versus host disease; therapy; cancer;
 KM organ transplant rejection disease; lymphoma; pancreatic disease;
 KM autoimmune disease; inflammatory disease; arthritis; binding site; ss.

XX Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS 1..447
 FT /*tag= a
 FT /note= "no stop codon given"

XX WO9945031-A2.
 XX
 PD 10-SEP-1999.
 XX
 PF 03-MAR-1999; 99WO-US04583.
 XX
 PR 03-MAR-1998; 98US-0034607.
 PR 03-FEB-1999; 99US-0244253.
 XX

XX (ABGE-) ABGENIX INC.

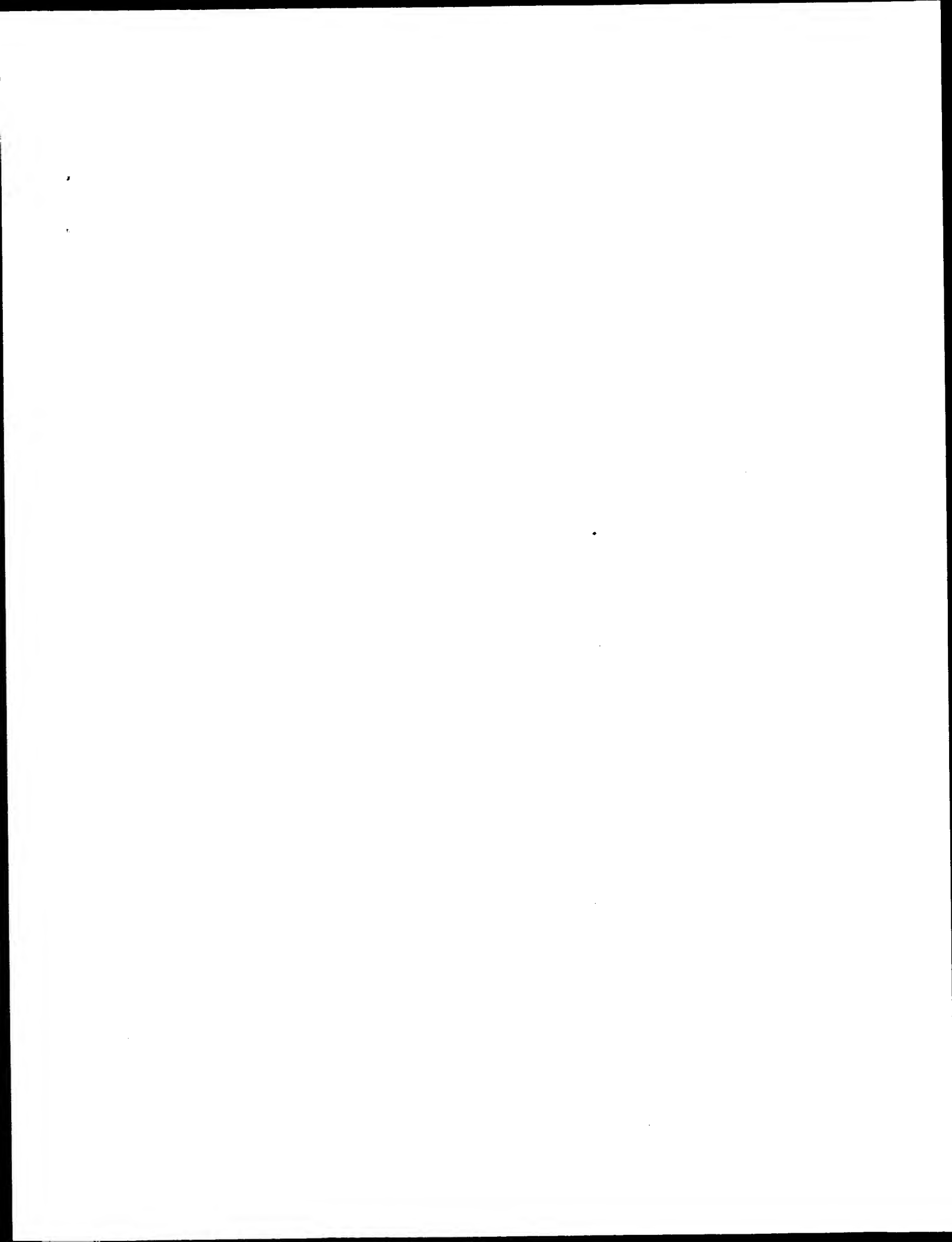
XX Davis CG, Blacher RM, Corvatan JR, Culwell AR, Green LT, Hales J;
 PI Havrilla N, Ivanov VE, Lipani JA, Liu Q, Weber RF, Yang X;
 XX
 DR WPI: 1999-540816/45.

DR P-PSDB; AAY34307.
 XX New monoclonal antibody, used for treating e.g. graft versus host
 PT disease, cancers, autoimmune diseases and inflammatory diseases
 XX
 PS Disclosure; Fig 30; 245pp; English.
 XX
 CC This sequence encodes the heavy chain of an antibody of the
 CC invention. The antibody is a monoclonal antibody (MAb) with an isotype
 CC that fixes complement and a variable region that binds to the epitope on
 CC CD147 bound by the IgM MAb ABX-CBL, providing that the antibody is not
 CC CBL1. The MAb can selectively kill activated T-cells, activated B-cells
 CC or resting or activated monocytes. The products and methods can be used
 CC for treating diseases involving activated T-cells or B-cells or
 CC monocytes, e.g. graft versus host disease (GVHD), organ transplant
 CC rejection diseases (e.g. renal transplant, ocular transplant), cancers
 CC (e.g. cancers of the blood (e.g. leukaemia's and lymphomas) and
 CC pancreatic), autoimmune diseases (e.g. lupus), and inflammatory diseases
 CC (e.g. arthritis).
 XX
 SQ Sequence 477 BP; 97 A; 157 C; 119 G; 104 T; 0 other;
 XX
 Query Match 52.0%; Score 13; DB 20; Length 477;
 Best Local Similarity 100.0%; Pred. No. 3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 6 aggtgcaggtcag 18
 |||||
 DB 13 AGGTGCAGCTCAG 1
 |||||
 RESULT 45
 AAV87894
 ID AAV87894 standard; cDNA; 527 BP.
 XX
 AC AAV87894;
 XX
 DT 12-FEB-1999 (first entry)
 XX
 DE EST clone FG380.
 XX
 KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
 KW tissue growth; actinin; inhibin; tumour invasion suppressor; EST; human;
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO9845437-A2.
 XX
 PD 15-OCT-1998.
 XX
 PF 10-APR-1998; 98WO-US06956.
 XX
 PR 10-APR-1997; 97US-0837312.
 XX
 PA (GEMV) GENETICS INST INC.
 XX
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Werberg D;
 PI Racie IA, Spaulding V, Treacy M;
 XX
 DR WPI; 1999-070078/06.
 XX
 PT New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries
 XX
 PS Claim 1; Page 215; 641pp; English.
 CC
 CC The present sequence represents an expressed sequence tag (EST), and is
 CC a polynucleotide of the invention. The polynucleotides of the invention
 CC are all secreted EST sequences isolated from a variety of human tissue

CC sources. The EST sequences and proteins encoded by them are predicted to
 CC have useful biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals, although no supporting data is given. Suggested activities
 CC include nutritional activity, immune stimulating or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC actinin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
 CC therapy.
 XX
 SQ Sequence 527 BP; 108 A; 162 C; 139 G; 118 T; 0 other;
 XX
 Query Match 52.0%; Score 13; DB 20; Length 527;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 7 ggtgcaggtcagc 19
 |||||
 DB 347 ggtgcaggtcagc 359
 |||||
 Search completed: October 9, 2001, 15:53:51
 Job time: 15192 sec

Wed Oct 10 07:45:36 2001

us-09-396-196f-10.oli.rng



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:55:42 ; Search time 218.82 Seconds
(without alignments)
21.629 Million cell updates/sec

Title: US-09-396-196f-10

Perfect score: 25
Sequence: 1 tctcagtgctcagtcagtcagcttg 25

Scoring table: OLIGO-NUC
Gapop 60.0 , Gapext 60.0

Searched: 324599 seqs, 9465562 residues

Word size : 9

Total number of hits satisfying chosen parameters: 12575

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCBUS.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	1041	2 US-08-401-068-7	Sequence 7, Appl
2	25	100.0	1041	2 US-08-846-338-7	Sequence 7, Appl
3	25	100.0	5872	3 US-08-411-768B-1	Sequence 1, Appl
4	25	100.0	5872	3 US-08-411-768B-6	Sequence 6, Appl
5	15	60.0	630	3 US-08-545-809A-10	Sequence 10, Appl
6	14	56.0	613	3 US-08-545-809A-5	Sequence 5, Appl
7	13	52.0	123	2 US-08-483-636-59	Sequence 59, Appl
8	13	52.0	123	2 US-08-483-632-59	Sequence 59, Appl
9	13	52.0	363	2 US-08-483-636-63	Sequence 63, Appl
10	13	52.0	363	2 US-08-483-632-63	Sequence 63, Appl
11	13	52.0	378	1 US-08-488-376-13	Sequence 13, Appl
12	13	52.0	378	1 US-08-488-376-15	Sequence 13, Appl
13	13	52.0	378	1 US-08-634-223-13	Sequence 13, Appl
14	13	52.0	378	2 US-08-634-223-15	Sequence 15, Appl
15	13	52.0	378	2 US-08-634-224-13	Sequence 13, Appl
16	13	52.0	378	2 US-08-634-224-15	Sequence 15, Appl
17	13	52.0	378	2 US-08-634-400-13	Sequence 13, Appl
18	13	52.0	378	2 US-08-634-400-15	Sequence 15, Appl
19	13	52.0	378	2 US-08-635-878-13	Sequence 13, Appl
20	13	52.0	378	2 US-08-635-878-15	Sequence 15, Appl
21	13	52.0	378	2 US-08-770-057-13	Sequence 13, Appl
22	13	52.0	378	2 US-08-770-057-15	Sequence 15, Appl
23	13	52.0	378	4 US-09-335-697B-13	Sequence 13, Appl
24	13	52.0	378	4 US-09-335-697B-15	Sequence 15, Appl
25	13	52.0	423	4 US-08-345-321-9	Sequence 9, Appl
26	13	52.0	546	3 US-08-545-809A-26	Sequence 26, Appl
27	13	52.0	718	3 US-09-154-083-31	Sequence 31, Appl

C 28	13	52.0	879	3 US-08-714-071-3	Sequence 3, Appl
C 29	13	52.0	1173	3 US-08-706-216-5	Sequence 5, Appl
C 30	13	52.0	1428	1 US-08-488-376-17	Sequence 17, Appl
C 31	13	52.0	1428	1 US-08-488-376-19	Sequence 19, Appl
C 32	13	52.0	1428	2 US-08-634-223-19	Sequence 19, Appl
C 33	13	52.0	1428	2 US-08-634-223-19	Sequence 19, Appl
C 34	13	52.0	1428	2 US-08-634-224-17	Sequence 17, Appl
C 35	13	52.0	1428	2 US-08-634-224-19	Sequence 19, Appl
C 36	13	52.0	1428	2 US-08-634-400-17	Sequence 17, Appl
C 37	13	52.0	1428	2 US-08-634-400-19	Sequence 19, Appl
C 38	13	52.0	1428	2 US-08-635-878-17	Sequence 17, Appl
C 39	13	52.0	1428	2 US-08-635-878-19	Sequence 19, Appl
C 40	13	52.0	1428	2 US-08-770-057-17	Sequence 17, Appl
C 41	13	52.0	1428	2 US-08-770-057-19	Sequence 19, Appl
C 42	13	52.0	1428	4 US-09-335-697B-17	Sequence 17, Appl
C 43	13	52.0	1428	4 US-09-335-697B-19	Sequence 19, Appl
C 44	13	52.0	2830	2 US-09-010-928B-1	Sequence 1, Appl
C 45	13	52.0	3105	4 US-08-542-635-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-401-068-7
Sequence 7, Application US/08401068
Patent No. 5859335
GENERAL INFORMATION:
APPLICANT: Patton, David
TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/401,068
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/351,970
FILING DATE: 08-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1038
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /product= "biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match	100.0%;	Score 25;	DB 2;	Length 1041;
Best Local Similarity	100.0%;	Pred. No. 5.6e-05;		
Matches 25; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

Oy	1	tccgtcagtlgcagtgcacgcttg	25
Db	111	TcGTCAgGTCAGGTCAGCAGCTTG	135

```

RESULT      2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO.5869719artis Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,338
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match      100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 5.0e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 tgcataagtcacagtcacacagctg 25
        |||||||
Db      111 tgcacagtcacagtcacacagcttg 135

RESULT      3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Bgrass
; APPLICANT: Martin Fuhrmann

```

```

1  APPLICANT: Nicholas Shaw
2  TITLE OF INVENTION: Biotechnological Method
3  TITLE OF INVENTION: of Producing Biotin
4  NUMBER OF SEQUENCES: 19
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
7  STREET: 30 Rockefeller Plaza
8  CITY: New York
9  STATE: New York
10 COUNTRY: USA
11 ZIP: 10112
12
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: Wordperfect
18 SOFTWARE: version 5.1
19
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: 05/08/411,768B
22 FILING DATE: 31-March-95
23 CLASSIFICATION: 435
24
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: CH 3124/92
27 FILING DATE: 02-OCT-1992
28
29 PRIOR APPLICATION DATA:
30 APPLICATION NUMBER: CH 2134/93
31 FILING DATE: 15-JUL-1993
32 INFORMATION FOR SEQ ID NO: 1:
33 SEQUENCE CHARACTERISTICS:
34 LENGTH: 5872 base pairs
35 TYPE: nucleic acid
36 STRANDEDNESS: double
37 TOPOLOGY: linear
38 MOLECULE TYPE: DNA (genomic)
39
40 HYPOTHETICAL: NO
41 ANTI-SENSE: NO
42 ORIGINAL SOURCE:
43 ORGANISM: Escherichia coli
44 STRAIN: DSM498
45 IMMEDIATE SOURCE:
46 CLONE: pBO30A-15/9
47
48 FEATURE:
49 NAME/KEY: CDS
50 LOCATION: 117..1157
51 IDENTIFICATION METHOD: experimental
52 OTHER INFORMATION: /codon_start=117
53 OTHER INFORMATION: /product="Biotin synthase"
54 OTHER INFORMATION: /evidence=EXPERIMENTAL
55 OTHER INFORMATION: /gene="biob"
56 OTHER INFORMATION: /number=1
57
58 FEATURE:
59 NAME/KEY: CDS
60 LOCATION: 2295..3050
61 OTHER INFORMATION: /codon_start=2295
62 OTHER INFORMATION: /function="involved in pimeloyl-CoA synthesis"
63 OTHER INFORMATION: /product="protein"
64 OTHER INFORMATION: /gene="bioc"
65 OTHER INFORMATION: /number=3
66
67 FEATURE:
68 NAME/KEY: CDS
69 LOCATION: 3750..5039
70 IDENTIFICATION METHOD: experimental
71 OTHER INFORMATION: /codon_start=3750
72 OTHER INFORMATION: /EC_number=2.6.1.62
73 OTHER INFORMATION: /product="DAPA synthase"
74 OTHER INFORMATION: /evidence=EXPERIMENTAL
75 OTHER INFORMATION: /gene="bioA"
76 OTHER INFORMATION: /number=5
77 OTHER INFORMATION: /standard_name="S-Adenosyl-L-methionine:8-amino-7-oxononanoate
78 OTHER INFORMATION: aminotransf."
79
80 NAME/KEY: CDS

```

LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 5098
OTHER INFORMATION: /function= "unknown", involved in biotin synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ORF1"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "promoter plac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name= "promoter plac"
FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name= "bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name= "bioa RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name= "ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name= "rho-independent"
OTHER INFORMATION: transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: 1..96
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "promoter plac"
OTHER INFORMATION: /evidence= EXPERIMENTAL
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 5e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tcgtcaggtcaggtcagcagtg 25
|||||
Db 227 TCGTCAGGTCAGGTCAGCAGGTTG 251

RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass

APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: Of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect
SOFTWARE: Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A15-9
FEATURE:
NAME/KEY: CDS
LOCATION: 1154..2308
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 1154
OTHER INFORMATION: /EC_number= 2.3.1.47
OTHER INFORMATION: /product= "KAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioF"
OTHER INFORMATION: /number= 2
FEATURE:
NAME/KEY: CDS
LOCATION: 3043..3753
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3043
OTHER INFORMATION: /EC_number= 6.3.3.3
OTHER INFORMATION: /product= "DTB synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biob"
OTHER INFORMATION: /number= 4
OTHER INFORMATION: /standard_name= "Dethiobiotin synthase"
FEATURE:
NAME/KEY: RBS
LOCATION: 1141..1156
OTHER INFORMATION: /standard_name= "biob RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3030..3045
OTHER INFORMATION: /standard_name= "biob RBS"
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1

FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-6

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 5e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tgcgcaggtcaggtcaggtcaggtg 25
|||||
Db 227 tgcgcaggtcaggtcaggtcaggtg 251

RESULT 5
US-08-545-809A-10/c
Sequence 10, Application US/08545809A
Patent No. 6096878
GENERAL INFORMATION:
APPLICANT: Honjo, Yasuko
APPLICANT: Matsuda, Fumihiko
TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE
TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
NUMBER OF SEQUENCES: 145
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545,809A
FILING DATE: 27-MAR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/00603
FILING DATE: 10-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John W.
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 06501/004001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 630 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL TYPE: human lymphoblast
CELL LINE: GM1
US-08-545-809A-10

Query Match 60.0%; Score 15; DB 3; Length 630;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtgcaggtcagca 20
|||||
Db 280 AGGTGCAGGTCA 266

RESULT 6
US-08-545-809A-5/c
Sequence 5, Application US/08545809A
Patent No. 6096878
GENERAL INFORMATION:
APPLICANT: Honjo, Yasuko
APPLICANT: Matsuda, Fumihiko
TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE
TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
NUMBER OF SEQUENCES: 145
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545,809A
FILING DATE: 27-MAR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/00603
FILING DATE: 10-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John W.
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 06501/004001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 613 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL TYPE: human lymphoblast
CELL LINE: GM1
US-08-545-809A-5

Query Match 56.0%; Score 14; DB 3; Length 613;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtgcaggtcagc 19
|||||
Db 283 AGGTGCAGGTCA 270

RESULT 7
US-08-483-636-59/c
Sequence 59, Application US/08483636
Patent No. 5914110
GENERAL INFORMATION:
APPLICANT: Holmes, Stephen D.
APPLICANT: Gross, Mitchell S.
APPLICANT: Sylvester, Daniel R.
TITLE OF INVENTION: Recombinant IL4 Antibodies Useful in
TITLE OF INVENTION: Treatment of IL4 Mediated Disorders
NUMBER OF SEQUENCES: 75
CORRESPONDENCE ADDRESS:
ADDRESSEE: Smitikline Beecham Corp./Corporate
ADDRESS: Intellectual Property

STREET: P.O. Box 1539 / UW2220
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,636
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/117366
FILING DATE: 07-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/136783
FILING DATE: 14-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US/94/10308
FILING DATE: 07-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Sutton, Jeffrey A.
REGISTRATION NUMBER: 34,028
REFERENCE/DOCKET NUMBER: P50186-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 270-5024
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 123 base pairs
TYPE: nucleic acid
STRANDEDNESS: Single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-483-636-59

Query Match 52.0%; Score 13; DB 2; Length 123;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtgcaggtcag 18
|||||

Db 70 AGGTGCAGGTGAG 58

RESULT 8
US-08-483-632-59/C
Sequence 59, Application US/08483632
Patent No. 5928904
GENERAL INFORMATION:
APPLICANT: Holmes, Stephen D.
APPLICANT: Gross, Mitchell S.
APPLICANT: Sylvester, Daniel R.
TITLE OF INVENTION: Recombinant IL4 Antibodies Useful in
TREATMENT OF ALLERGIC DISORDERS
NUMBER OF SEQUENCES: 75
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SmithKline Beecham Corp./Corporate
ADDRESS: Intellectual Property
STREET: P.O. Box 1539 / UW2220
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,632
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/117366
FILING DATE: 07-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/136783
FILING DATE: 14-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US/94/10308
FILING DATE: 07-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Sutton, Jeffrey A.
REGISTRATION NUMBER: 34,028
REFERENCE/DOCKET NUMBER: P50186-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 270-5024
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 123 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-483-632-59

Query Match 52.0%; Score 13; DB 2; Length 123;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtgcaggtcag 18
|||||

Db 70 AGGTGCAGGTGAG 58

RESULT 9
US-08-483-636-63/C
Sequence 63, Application US/08483636
Patent No. 5914110
GENERAL INFORMATION:
APPLICANT: Holmes, Stephen D.
APPLICANT: Gross, Mitchell S.
APPLICANT: Sylvester, Daniel R.
TITLE OF INVENTION: Recombinant IL4 Antibodies Useful in
TREATMENT OF ALLERGIC DISORDERS
NUMBER OF SEQUENCES: 75
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SmithKline Beecham Corp./Corporate
ADDRESS: Intellectual Property
STREET: P.O. Box 1539 / UW2220
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,636
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/117366
FILING DATE: 07-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/136783
FILING DATE: 14-OCT-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US/94/10308
FILING DATE: 07-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Sulton, Jeffrey A.
REGISTRATION NUMBER: 34,028
REFERENCE/DOCKET NUMBER: P50186-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 270-5024
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 363 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-483-636-63

Query Match 52.0%; Score 13; DB 2; Length 363;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtgcaggtcag 18
|||||
DB 70 AGGTGCAGGTCTCAG 58

RESULT 10
US-08-483-632-63/c
Sequence 63, Application US/08483632
Patent No. 5928904
GENERAL INFORMATION:
APPLICANT: Holmes, Stephen D.
APPLICANT: Gross, Mitchell S.
APPLICANT: Sylvester, Daniel R.
TITLE OF INVENTION: Recombinant IL4 Antibodies Useful in
TITLE OF INVENTION: Treatment of IL4 Mediated Disorders
NUMBER OF SEQUENCES: 75
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corp./Corporate
ADDRESSEE: Intellectual Property
STREET: P.O. Box 1539 / UM2220
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,632
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/117366
FILING DATE: 07-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/136783
FILING DATE: 14-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US/94/10308
FILING DATE: 07-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Sulton, Jeffrey A.
REGISTRATION NUMBER: 34,028
REFERENCE/DOCKET NUMBER: P50186-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 270-5024
TELEFAX: (215) 270-5090

INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 363 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-483-632-63

Query Match 52.0%; Score 13; DB 2; Length 363;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtgcaggtcag 18
|||||
DB 70 AGGTGCAGGTCTCAG 58

RESULT 11
US-08-488-376-13/c
Sequence 13, Application US/08488376
Patent No. 5811524
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,376
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-488-376-13

Query Match 52.0%; Score 13; DB 1; Length 378;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 ggtgcaggtcagc 19
|||||
Db 69 GGTGCAGGTGACG 57

RESULT 12

US-08-488-376-15/c

; Sequence 15, Application US/08488376
; Patent No. 581524

GENERAL INFORMATION:

APPLICANT: BRAMS, Peter

APPLICANT: CHAMAT, Soulatima Salim

APPLICANT: PAN, Li-Zhen

APPLICANT: WALSH, Edward E.

APPLICANT: HEARD, Cheryl Janne

APPLICANT: NEWMAN, Roland Anthony

TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN

TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND

TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESS: Burns, Doane, Swecker & Mathis

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/488,376

FILING DATE: 07-JUN-1995

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Teskin, Robin L.

REGISTRATION NUMBER: 35,030

REFERENCE/DOCKET NUMBER: 012712-150

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 378 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 1..378

US-08-488-376-15

Query Match

Best local Similarity 52.0%; Score 13; DB 1; Length 378;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 aggtgcaggtcagc 18
|||||
Db 70 AGGTGCAGGTGACG 58

RESULT 13

US-08-634-223-13/c

; Sequence 13, Application US/08634223

; Patent No. 5840298

; GENERAL INFORMATION:

APPLICANT: BRAMS, Peter

APPLICANT: CHAMAT, Soulatima Salim

APPLICANT: PAN, Li-Zhen

APPLICANT: WALSH, Edward E.

APPLICANT: HEARD, Cheryl Janne

APPLICANT: NEWMAN, Roland Anthony

TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN

TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND

TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESS: Burns, Doane, Swecker & Mathis

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/634,223

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/488,376

FILING DATE: 07-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Teskin, Robin L.

REGISTRATION NUMBER: 35,030

REFERENCE/DOCKET NUMBER: 012712-150

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 378 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 1..378

US-08-634-223-13

Query Match

Best local Similarity 52.0%; Score 13; DB 2; Length 378;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 ggtgcaggtcagc 19
|||||
Db 69 GGTGCAGGTGACG 57

RESULT 14

US-08-634-223-15/c

; Sequence 15, Application US/08634223

; Patent No. 5840298

; GENERAL INFORMATION:

APPLICANT: BRAMS, Peter

APPLICANT: CHAMAT, Soulatima Salim

APPLICANT: PAN, Li-Zhen

APPLICANT: WALSH, Edward E.

APPLICANT: HEARD, Cheryl Janne

APPLICANT: NEWMAN, Roland Anthony

TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN

TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND

TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESS: Burns, Doane, Swecker & Mathis

STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,223
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-634-223-15

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtcaggtcag 18
|||||
Db 70 Aggtcaggtcag 58

RESULT 15
US-08-634-224-13/C
Sequence 13, Application US/08634224
Patent No. 5866125
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,224
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-634-224-13

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ggtcaggtcagc 19
|||||
Db 69 Ggtcaggtcagc 57

RESULT 16
US-08-634-224-15/C
Sequence 15, Application US/08634224
Patent No. 5866125
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,224
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030

REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-634-224-15

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 aggtgcagtcacg 18
|||||
DB 70 AGGTGCAGTCACG 58

RESULT 17
US-08-634-400-13/c
Sequence 13, Application US/08634400
Patent No. 5939068
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,400
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-634-400-13

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 ggtgcagtcacg 19
|||||
DB 69 GGTGCAGTCACG 57

RESULT 18
US-08-634-400-15/c
Sequence 15, Application US/08634400
Patent No. 5939068
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,400
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-634-400-15

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 aggtgcaggtcag 18
|||||
Db 70 AGGTGCAGGTGAC 58

RESULT 19
US-08-635-878-13/c

; Sequence 13, Application US/08635878
; Patent No. 595364
; GENERAL INFORMATION:
; APPLICANT: BRAMS, Peter
; APPLICANT: CHAMAT, Soulaïma Salim
; APPLICANT: PAN, Li-Zhen
; APPLICANT: WALSH, Edward E.
; APPLICANT: HEARD, Cheryl Janne
; APPLICANT: NEWMAN, Roland Anthony
; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
; TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/635,878
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/488,376
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 012712-150
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 378 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..378
; US-08-635-878-13

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 82;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 ggtgcaggtcagc 19
|||||
Db 69 GGTGCAGGTGACG 57

RESULT 20
US-08-635-878-15/c
; Sequence 15, Application US/08635878
; Patent No. 595364

; GENERAL INFORMATION:
; APPLICANT: BRAMS, Peter
; APPLICANT: CHAMAT, Soulaïma Salim
; APPLICANT: PAN, Li-Zhen
; APPLICANT: WALSH, Edward E.
; APPLICANT: HEARD, Cheryl Janne
; APPLICANT: NEWMAN, Roland Anthony
; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
; TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/635,878
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/488,376
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 012712-150
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 378 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..378
; US-08-635-878-15

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 82;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 aggtgcaggtcag 18
|||||
Db 70 AGGTGCAGGTGAC 58

RESULT 21

US-08-770-057-13/c
; Sequence 13, Application US/08770057
; Patent No. 5958765
; GENERAL INFORMATION:
; APPLICANT: BRAMS, Peter
; APPLICANT: CHAMAT, Soulaïma Salim
; APPLICANT: PAN, Li-Zhen
; APPLICANT: WALSH, Edward E.
; APPLICANT: HEARD, Cheryl Janne
; APPLICANT: NEWMAN, Roland Anthony
; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
; TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF

NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,057
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-770-057-13

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 7 ggtgcagtcagc 19
|||||
Db 69 GGTGCAGTCAGC 57

RESULT 22
US-08-770-057-15/c
Sequence 15, Application US/08770057
Patent No. 5958765
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,057
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
US-08-770-057-15

Query Match 52.0%; Score 13; DB 2; Length 378;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 aggtgcagtcagc 18
|||||
Db 70 AGGTGCAGTCAGC 58

RESULT 23
US-09-335-697B-13/c
Sequence 13, Application US/09335697B
Patent No. 6200804
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/335,697B
FILING DATE: 06-Jul-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/770,057
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-335-697B-13

Query Match 52.0%; Score 13; DB 4; Length 378;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 ggtgcaggtcagc 19
|||||
DB 69 ggtgcaggtcagc 57

RESULT 24
US-09-335-697B-15/C
Sequence 15, Application US/09335697B
Patent No. 6200804
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
CHAMAT, Soulatima Salim
PAN, Li-Zhen
WALSH, Edward E.
HEARD, Cheryl Janne
NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/335,697B
FILING DATE: 06-Jul-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/770,057
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:

LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..378
SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-09-335-697B-15

Query Match 52.0%; Score 13; DB 4; Length 378;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 aggtgcaggtcag 18
|||||
DB 70 aggtgcaggtcag 58

RESULT 25
US-08-345-321-9/C
Sequence 9, Application US/08345321
Patent No. 5914109

GENERAL INFORMATION:
APPLICANT: ZOLLA-PAZNER, Susan
GORN, Matrosav K.
TITLE OF INVENTION: HETEROHYBRIDOMAS PRODUCING HUMAN
MONOCLONAL ANTIBODIES TO HIV-1
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Broadway and Neimark
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/345,321
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/872,675
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Broadway, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: ZOLLA-PAZNER1B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..423
US-08-345-321-9

Query Match 52.0%; Score 13; DB 2; Length 423;
Best Local Similarity 100.0%; Pred. No. 81;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 aggtgcagtcag 18
|||||

Db 127 AGGTGCAGTCAG 115

RESULT 26

US-08-545-809A-26/C

Sequence 26, Application US/08545809A

Patent No. 6096878

GENERAL INFORMATION:

APPLICANT: Honjo, Tasuku

APPLICANT: Matsuda, Fumihiko

TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE

TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME

NUMBER OF SEQUENCES: 145

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson, P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: US

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/545,809A

FILING DATE: 27-MAR-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP93/00603

FILING DATE: 10-MAY-1993

ATTORNEY/AGENT INFORMATION:

NAME: Freeman, John W.

REGISTRATION NUMBER: 29,066

REFERENCE/DOCKET NUMBER: 06501/004001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-542-5070

TELEFAX: 617-542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 546 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

CELL TYPE: human lymphoblast

CELL LINE: CGM1

US-08-545-809A-26

Query Match 52.0%; Score 13; DB 3; Length 546;

Best Local Similarity 100.0%; Pred. No. 80;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 ggtcagtcagc 19
|||||

Db 232 GGTCCAGTCAGC 220

RESULT 27

US-09-154-083-31

Sequence 31, Application US/09154083

Patent No. 6150513

GENERAL INFORMATION:

APPLICANT: Wu, Kai

TITLE OF INVENTION: Polyketide Synthase Enzymes and Recombinant DNA

TITLE OF INVENTION: Constructs Therefor

FILE REFERENCE: 30062-20022.00 Polyketide Synthase Enz

CURRENT APPLICATION NUMBER: US/09/154,083

CURRENT FILING DATE: 1998-09-16

NUMBER OF SEQ ID NOS: 31

SOFTWARE: Patentln Ver. 2.0

SEQ ID NO 31

LENGTH: 718

TYPE: DNA

ORGANISM: Streptomyces hygroscopicus

US-09-154-083-31

Query Match 52.0%; Score 13; DB 3; Length 718;

Best Local Similarity 100.0%; Pred. No. 79;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gcagtcagcagc 22
|||||

Db 347 gcagtcagcagc 359

RESULT 28

US-08-714-071-3/C

Sequence 3, Application US/08714071

Patent No. 6136584

GENERAL INFORMATION:

APPLICANT: Tsutomu, FUJIMURA

APPLICANT: Shiro, OKUNO

APPLICANT: Hisanobu, HIRANO

APPLICANT: Sadahito, SHIN

TITLE OF INVENTION: FK506 BINDING PROTEIN GENE

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sughrue, Mion, Zimm, Macpeak & Seas

STREET: 2100 Pennsylvania Avenue, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: United States

ZIP: 20037-3202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/714,071

FILING DATE:

CLASSIFICATION: 435

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 293-7060

TELEFAX: (202) 293-7860

TELEX: 6491103

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 879 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Human fetal brain cDNA library

IMMEDIATE SOURCE:

CLONE: OTK4(6-1)

FEATURE:

NAME/KEY: CDS

LOCATION: 70..393

US-08-714-071-3

Query Match 52.0%; Score 13; DB 3; Length 879;

Best Local Similarity 100.0%; Pred. No. 78;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ggtgcaggtcagc 19
|||||
Db 303 GGTGCAGGTCTCAGC 291

RESULT 29

US-08-706-216-5/c
; Sequence 5, Application US/08706216
; Patent No. 6140098
; GENERAL INFORMATION:
; APPLICANT: Balasubramanian, SriRam
; APPLICANT: Ford, John
; APPLICANT: Gorman, Daniel M.
; APPLICANT: Zurawski, Gerard
; TITLE OF INVENTION: MAMMALIAN PROTEASES; RELATED REAGENTS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/706,216
; FILING DATE: 30-AUG-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0613
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-852-9196
; TELEFAX: 415-496-1200
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1173 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1173
; US-08-706-216-5

Query Match 52.0%; Score 13; DB 3; Length 1173;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcaggtgcaggtc 16
|||||
Db 418 TCAGGTGCAGGTCTC 406

RESULT 30

US-08-488-376-17/c
; Sequence 17, Application US/08488376
; Patent No. 5811524
; GENERAL INFORMATION:
; APPLICANT: BRAMS, Peter
; APPLICANT: CHAMAT, Soulaïma Salim
; APPLICANT: PAN, Li-Zhen
; APPLICANT: WALSH, Edward E.

APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,376
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-488-376-17

Query Match 52.0%; Score 13; DB 1; Length 1428;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ggtgcaggtcagc 19
|||||
Db 126 GGTGCAGGTCTCAGC 114

RESULT 31

US-08-488-376-19/c
; Sequence 19, Application US/08488376
; Patent No. 5811524
; GENERAL INFORMATION:
; APPLICANT: BRAMS, Peter
; APPLICANT: CHAMAT, Soulaïma Salim
; APPLICANT: PAN, Li-Zhen
; APPLICANT: WALSH, Edward E.
; APPLICANT: HEARD, Cheryl Janne
; APPLICANT: NEWMAN, Roland Anthony
; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,376
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-488-376-19

Query Match 52.0%; Score 13; DB 1; Length 1428;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 aggtgcaggtcag 18
|||||
DB 127 AGTCGAGCTCAG 115
RESULT 32
US-08-634-223-17/c
Sequence 17, Application US/08634223
Patent No. 5840298
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,223
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:

NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-634-223-17

Query Match 52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 ggtgcaggtcagc 19
|||||
DB 126 GTTCGAGCTCAGC 114

RESULT 33
US-08-634-223-19/c
Sequence 19, Application US/08634223
Patent No. 5840298
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,223
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-634-223-19

Query Match 52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtgcaggtcag 18
|||||
Db 127 AGGTGCAGGTCTAG 115

RESULT 34

US-08-634-224-17/c
Sequence 17, Application US/08634224
Patent No. 5866125

GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,224
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-634-224-17

Query Match 52.0%; Score 13; DB 2; Length 1428;

Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 ggtgcaggtcagc 19
|||||
Db 126 GTGTGAGGTCTAG 114

RESULT 35

US-08-634-224-19/c
Sequence 19, Application US/08634224
Patent No. 5866125

GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/634,224
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-634-224-19

Query Match 52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtgcaggtcag 18
|||||
Db 127 AGGTGCAGGTCTAG 115

RESULT 36
US-08-634-400-17/c

```
; Sequence 17, Application US/08634400
; Patent No. 5938068
; GENERAL INFORMATION:
; APPLICANT: BRAMS, Peter
; APPLICANT: CHAMAT, Soulaïma Salim
; APPLICANT: PAN, Li-Zhen
; APPLICANT: WALSH, Edward E.
; APPLICANT: HEARD, Cheryl Janne
; APPLICANT: NEWMAN, Roland Anthony
; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
; TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/634,400
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/488,376
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 012712-150
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1428 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1428
; US-08-634-400-17

Query Match          52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      7 ggtgcaggtcagc 19
        |||
Db      126 ggtgcaggtcagc 114

RESULT 37
US-08-634-400-19/c
; Sequence 19, Application US/08634400
; Patent No. 5939068
; GENERAL INFORMATION:
; APPLICANT: BRAMS, Peter
; APPLICANT: CHAMAT, Soulaïma Salim
; APPLICANT: PAN, Li-Zhen
; APPLICANT: WALSH, Edward E.
; APPLICANT: HEARD, Cheryl Janne
; APPLICANT: NEWMAN, Roland Anthony
; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
```

```
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
; TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/634,400
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/488,376
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 012712-150
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1428 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1428
; US-08-634-400-19

Query Match          52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      6 aggtgcaggtcag 18
        |||
Db      127 aggtgcaggtcag 115

RESULT 38
US-08-635-878-17/c
; Sequence 17, Application US/08635878
; Patent No. 5955364
; GENERAL INFORMATION:
; APPLICANT: BRAMS, Peter
; APPLICANT: CHAMAT, Soulaïma Salim
; APPLICANT: PAN, Li-Zhen
; APPLICANT: WALSH, Edward E.
; APPLICANT: HEARD, Cheryl Janne
; APPLICANT: NEWMAN, Roland Anthony
; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
; TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/635,878
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-635-878-17

Query Match      52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 ggtgcaggtcagc 19
Db      126 GGTGCAGGTCTCAGC 114

RESULT 39
US-08-635-878-19/c
Sequence 19, Application US/08635878
Patent No. 5955364
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/635,878
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
```

```
APPLICATION NUMBER: 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-635-878-19

Query Match      52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 aggtgcaggtcag 18
Db      127 AGGTGCAGGTCTCAG 115

RESULT 40
US-08-770-057-17/c
Sequence 17, Application US/08770057
Patent No. 5958765
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
TITLE OF INVENTION: METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,057
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 17:
```

SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-770-057-17

Query Match 52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ggtcagatcagc 19
|||||
DB 126 GGTGACAGTCAGC 114

RESULT 41
US-08-770-057-19/c
Sequence 19, Application US/08770057
Patent No. 5958765
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,057
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/488,376
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-08-770-057-19

Query Match 52.0%; Score 13; DB 2; Length 1428;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 aggtcagatcag 18
|||||
DB 127 AGGTGACAGTCAGC 115

RESULT 42
US-09-335-697B-17/c
Sequence 17, Application US/09335697B
Patent No. 6200804
GENERAL INFORMATION:
APPLICANT: BRAMS, Peter
APPLICANT: CHAMAT, Soulaïma Salim
APPLICANT: PAN, Li-Zhen
APPLICANT: WALSH, Edward E.
APPLICANT: HEARD, Cheryl Janne
APPLICANT: NEWMAN, Roland Anthony
TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/335,697B
FILING DATE: 06-Jul-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/770,057
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1428 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1428
US-09-335-697B-17

Query Match 52.0%; Score 13; DB 4; Length 1428;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ggtcagatcagc 19
|||||
DB 126 GGTGACAGTCAGC 114

```

RESULT 43
US-09-335-697B-19/c
; Sequence 19, Application US/09335697B
; Patent No. 6200804
; GENERAL INFORMATION:
; APPLICANT: BRAMS, Peter
; CHAMAT, Soulaïma Salim
; PAN, Li-Zhen
; WALSH, Edward E.
; HEARD, Cheryl Janne
; NEWMAN, Roland Anthony
; TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
; MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
; METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/335,697B
; FILING DATE: 06-Jul-2000
; CLASSIFICATION: <Unknown>
; PRIORITY INFORMATION DATA:
; APPLICATION NUMBER: 08/770,057
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 012712-150
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1428 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1428
; SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-335-697B-19

Query Match 52.0%; Score 13; DB 4; Length 1428;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 aggtgcaggtcag 18
|||||
Db 127 AGGTGCAGGTcAG 115

RESULT 44
US-09-010-928B-1
; Sequence 1, Application US/09010928B
; Patent No. 5994099
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V
; APPLICANT: Hayashi, Cheryl Y

```

```

; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 CATHOUSE RD. SUITE 500E
; CITY: FALLS CHURCH
; STATE: VIRGINIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010,928B
; FILING DATE: 22-JAN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28977
; REFERENCE/DOCKET NUMBER: 1447-109P
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2830 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..2830
; OTHER INFORMATION: /note="Flagelliform DNA sequence
; OTHER INFORMATION: taken from the 5' region. The putative start codon is at
; OTHER INFORMATION: position 219"
; NAME/KEY: CDS
; LOCATION: 219..2830
US-09-010-928B-1

Query Match 52.0%; Score 13; DB 2; Length 2830;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tcaggtcaggtc 16
|||||
Db 588 TCAGGTcAGGTc 600

RESULT 45
US-08-542-635-1
; Sequence 1, Application US/08542635
; Patent No. 6218356
; GENERAL INFORMATION:
; APPLICANT: Pawson, Anthony
; APPLICANT: Henkemeyer, Mark
; APPLICANT: Letwin, Kenneth
; TITLE OF INVENTION: NOVEL NEURAL RECEPTOR
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bereskin & Parr
; STREET: 40 King Street West, Box 401
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```


OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/542,635
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Mcdiarmid, Shona S.
REGISTRATION NUMBER: 38,798
REFERENCE/DOCKET NUMBER: 3153-162
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 364-7311
TELEFAX: (416) 361-1398
TELEX: 06-23115
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3105 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Mus musculus
DEVELOPMENTAL STAGE: Embryo
IMMEDIATE SOURCE:
LIBRARY: lambda gt10 cDNA library
CLONE: Combined PUKRACE A2 and K2 AND cDNA clones
POSITION IN GENOME:
CHROMOSOME/SEGMENT: Distal end of chromosome 4
MAP POSITION: near the and-1 mutation
US-08-542-635-1

Query Match 52.0%; Score 13; DB 4; Length 3105;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 tgcaggtcagcac 21
|||||
Db 3061 TGCAGGTCAGCAC 3073

Search completed: October 9, 2001, 15:55:43
Job time: 13219 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:15:39 ; Search time 5323.87 Seconds
(without alignments)
44.389 Million cell updates/sec

Title: US-09-396-196f-10

Perfect score: 25

Sequence: 1 tcgtcagtgatgcagtcagcagcttg 25

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 9

Total number of hits satisfying chosen parameters: 513911

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

```

117: gb_est148:*
118: gb_est149:*
119: gb_est150:*
120: gb_est151:*
121: gb_est152:*
122: gb_est153:*
123: gb_est154:*
124: gb_est155:*
125: gb_est156:*
126: gb_est157:*
127: gb_est158:*
128: gb_est159:*
129: gb_est160:*
130: gb_est161:*
131: gb_est162:*
132: gb_est163:*
133: gb_est164:*
134: gb_est165:*
135: gb_est166:*
136: gb_est167:*
137: gb_est168:*
138: gb_est169:*
139: gb_est170:*
140: gb_est171:*
141: gb_est172:*
142: gb_est173:*
143: gb_est174:*
144: gb_est175:*
145: gb_est176:*
146: gb_est177:*
147: gb_est178:*
148: gb_est179:*
149: gb_est180:*
150: gb_est181:*
151: gb_est182:*
152: gb_est183:*
153: gb_est184:*
154: gb_est185:*
155: gb_est186:*
156: gb_est187:*
157: gb_est188:*
158: gb_est189:*
159: gb_est190:*
160: gb_est191:*
161: gb_est192:*
162: gb_est193:*
163: gb_est194:*
164: gb_est195:*
165: gb_est196:*
166: gb_est197:*
167: gb_est198:*
168: gb_est199:*
169: gb_est200:*
170: gb_est201:*
171: gb_est202:*
172: gb_est203:*
173: gb_est204:*
174: gb_est205:*
175: gb_est206:*
176: gb_est207:*
177: gb_est208:*
178: gb_est209:*
179: gb_est210:*
180: gb_est211:*
181: gb_est212:*
182: gb_est213:*
183: gb_est214:*
184: gb_est215:*
185: gb_est216:*
186: gb_est217:*
187: gb_est218:*
188: gb_est219:*
189: gb_est220:*

```

```

190: gb_est110:*
191: gb_est111:*
192: gb_est112:*
193: gb_est113:*
194: gb_est114:*
195: gb_est115:*
196: gb_est116:*
197: gb_est117:*
198: gb_est118:*
199: gb_est119:*
200: gb_est120:*
201: gb_est121:*
202: gb_est122:*
203: gb_est123:*
204: gb_est124:*
205: gb_est125:*
206: gb_est126:*
207: gb_est127:*
208: gb_est128:*
209: gb_est129:*
210: gb_est130:*
211: gb_est131:*
212: gb_est132:*
213: gb_est133:*
214: gb_est134:*
215: gb_est135:*
216: gb_est136:*
217: gb_est137:*
218: gb_est138:*
219: gb_est139:*
220: gb_est140:*
221: gb_est141:*
222: gb_est142:*
223: gb_est143:*
224: gb_est144:*
225: gb_est145:*
226: gb_est146:*
227: gb_est147:*
228: gb_est148:*
229: gb_est149:*
230: gb_est150:*
231: gb_est151:*
232: gb_est152:*
233: gb_est153:*
234: gb_est154:*
235: gb_est155:*
236: gb_est156:*
237: gb_est157:*
238: gb_est158:*
239: gb_est159:*
240: gb_est160:*
241: gb_est161:*
242: gb_est162:*
243: gb_est163:*
244: gb_est164:*
245: gb_est165:*
246: gb_est166:*
247: gb_est167:*
248: gb_est168:*
249: gb_est169:*
250: gb_est170:*
251: gb_est171:*
252: gb_est172:*
253: gb_est173:*
254: gb_est174:*
255: gb_est175:*
256: gb_est176:*
257: gb_est177:*
258: gb_est178:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	17	68.0	324	149	BF464374 UI-M-CG0P
C 2	17	68.0	413	141	BE862745 UI-M-BH0-
C 3	17	68.0	495	24	A1789420 UK99a06.y
C 4	17	68.0	641	243	AZ426129 1M0206H03
C 5	16	64.0	163	1	AA065941 m151a01.r
C 6	16	64.0	288	160	BB516544 BB516544
C 7	16	64.0	328	190	W54143 md13c03.r1
C 8	16	64.0	338	190	W34110 mc57c06.r1
C 9	16	64.0	345	31	AV633196 AV633196
C 10	16	64.0	368	2	AA110497 m162e03.r
C 11	16	64.0	370	121	AM822106 uq13c09.x
C 12	16	64.0	395	148	BF444068 261798 MA
C 13	16	64.0	398	226	AQ298728 HS_3110.B
C 14	16	64.0	430	20	A1467556 ve36g09.x
C 15	16	64.0	544	122	AM892602 CM3-NN000
C 16	16	64.0	553	229	AO530462 RPEC1-11-3
C 17	16	64.0	775	136	BE535032 601233227
C 18	16	64.0	822	77	BE194624 HVSMEH008
C 19	16	64.0	897	220	AL176719 Tetradodon
C 20	16	64.0	976	174	BG172968 602336627
C 21	16	64.0	1011	192	AK010837 MUS muscu
C 22	16	64.0	1298	152	BG328084 602427134
C 23	15	60.0	180	10	AA699872 zj81f04.s
C 24	15	60.0	264	157	H22139 y138a03.s1
C 25	15	60.0	279	187	R50061 y159c10.s1
C 26	15	60.0	292	19	AT331540 fa94b01.y
C 27	15	60.0	303	251	AA877721 RPEC1-23-1
C 28	15	60.0	300	127	BB175346 BB175346
C 29	15	60.0	306	173	BG100182 ux94g04.y
C 30	15	60.0	382	16	AA1138808 qd98g02.x
C 31	15	60.0	385	117	AM522699 UI-R-B00-
C 32	15	60.0	396	116	AA486040 69428 MAR
C 33	15	60.0	407	1	AA062258 m155a08.r
C 34	15	60.0	407	151	BF599674 BF599674
C 35	15	60.0	408	224	AO085001 HS_2265_A
C 36	15	60.0	409	151	BF653430 AA161616
C 37	15	60.0	416	2	AA146146 m665d12.r
C 38	15	60.0	446	13	AA881231 vx10d04.r
C 39	15	60.0	457	151	BF604067 267755 MA
C 40	15	60.0	477	235	AO921326 RPEC1-23-2
C 41	15	60.0	492	140	BE808281 213447 MA
C 42	15	60.0	495	240	AZ4261270 RPEC1-23-4
C 43	15	60.0	496	151	BF652819 276504 MA
C 44	15	60.0	525	150	BF567735 UI-R-B00-
C 45	15	60.0	529	114	AM290780 NXNV046H0

ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL
1	BF464374/c	324 bp mRNA	EST	04-DEC-2000						
	UI-M-CG0P-bog-f-09-0-UI-s1	NIH_BMAP_Ret4_S2	Mus musculus	CDNA clone						
	UI-M-CG0P-bog-f-09-0-UI-3'	mRNA sequence.								
	BF464374									
	BF464374.1	GI:11533557								
	EST.									
	house mouse.									
	Mus musculus									
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;									
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.									
	1 (bases 1 to 324)									
	Bonaldo,M.F., Lennon,G. and Soares,M.B.									
	Normalization and subtraction: two approaches to facilitate gene									
	discovery									
	Genome Res. 6 (9), 791-806 (1996)									

MEDLINE

97044477
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mestr@mail.nih.gov

COMMENT

Oligo-dT track not found, Not I site shown in beginning of sequence is likely internal to the message. CDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP CDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements. The following repetitive elements were found in this cDNA sequence: 102-169, >(GAA)n#Simple-repeat
Seq primer: M13 Forward
POLYA-NO.

FEATURES

SOURCE

Location/Qualifiers
1..324
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-CG0P-bog-f-09-0-UI"
/clone_1lb="NIH_BMAP_Ret4_S2"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7SD-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_Ret4_S2 library is a subtracted library, ultimately derived from mouse retina tissue libraries at various stages of development. For a detailed description of the library from which this clone was derived, please visit our web site at brainest.eng.uiowa.edu.
TAG_SEQ=None found"

BASE COUNT

91 a 84 c 89 g 60 t

Query Match 68.0%; Score 17; DB 149; Length 324;

Best Local Similarity 100.0%; Pred. No. 21;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cgtcaggtgcaggtcag 18

DB 324 CGTCAGGTGCAGGTGACG 308

RESULT 2 BE862745 413 bp mRNA EST 29-SEP-2000

LOCUS UI-M-BH0-ej-f-b-11-0-UI.r1 NIH_BMAP_M_S1 Mus musculus CDNA clone

DEFINITION UI-M-BH0-ej-f-b-11-0-UI 5', mRNA sequence.

ACCESSION BE862745

VERSION BE862745.1 GI:10382015

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 413)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

CONTACT Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: mestr@mail.nih.gov

/clone_11b="Mouse 10kb plasmid UUC1M library"
 /sex="male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g1147321149b/AP12972.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT
 156 a 184 c 183 g 118 t

ORIGIN

Query Match 68.0%; Score 17; DB 243; Length 641;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 cgtcagtgacagtcag 18
 ||||||||||||||||
 Db 212 CGTCAGTGACAGTCAG 196

RESULT 5
 AA065941/c 163 bp mRNA EST 03-FEB-1997
 LOCUS m15101.r1 Stratagene mouse testis (#937308) Mus musculus CDNA
 DEFINITION clone IMAGE:515496 5', mRNA sequence.
 ACCESSION AA065941
 VERSION AA065941.1 GI:1562644
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 163)
 Author(s): Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
 TITLE The WashU-HMNI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Marra M/Mouse EST Project
 WashU-HMNI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.wustl.edu
 This clone is available royalty-free through INL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:09344
 Seq primer: -28m13 rev1 ET from Amerham
 High quality sequence stop: 144.
 Location/Qualifiers
 1. 163
 /organism="Mus musculus"
 /strain="Inbred CD-1"
 /db_xref="taxon:10090"
 /clone="IMAGE:515496"

/clone_11b="Stratagene mouse testis (#937308)"
 /sex="males"
 /tissue_type="testis"
 /dev_stage="10-12 week old"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: testis; Vector: pBluescript SK-; Site: 1: EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer: 01190 dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGACGAG 3' -3' adaptor sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3'."

BASE COUNT
 23 a 56 c 49 g 35 t

ORIGIN

Query Match 64.0%; Score 16; DB 1; Length 163;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 gtcagtgacagtcag 18
 ||||||||||||||||
 Db 16 GTcAGTGACAGTCAG 1

RESULT 6
 BB516544/c 288 bp mRNA EST 28-JUL-2000
 LOCUS BB516544 RIKEN full-length enriched, 16 days neonate heart Mus
 DEFINITION musculus cDNA clone D830016711 3' similar to M76601 Mouse alpha
 cardiac myosin heavy chain mRNA, mRNA sequence.
 ACCESSION BB516544
 VERSION BB516544.1 GI:9568002
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 288)
 Author(s): Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigenoto, Y., Shimagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watanabe, S., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
 TITLE RIKEN Mouse ESTs (Kono, H., et al.)
 JOURNAL Unpublished (2000)
 COMMENT Contact: Yoshihide Hayashizaki
 Genome Exploration Research Group, Life Science Tsukuba Center,
 The Institute of Physical and Chemical Research (RIKEN), Genomic
 Sciences Center
 3-1-1 Koyada, Tsukuba, Ibaraki 305-0074, Japan
 Tel: +81-298-36-9013
 Fax: +81-298-36-9098
 Email: genome-rtc@ric.riken.go.jp/
 URL: http://genome-rtc.riken.go.jp/
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagao, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Thermostabilization and thermoactivation of the thermostable enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh, M., Katsunari, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, T., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.
Location/Qualifiers
1. .288
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="DB30016j11"
/clone_1ib="RIKEN full-length enriched, 16 days neonate heart"
/tissue-type="heart"
/dev_stage="16 days neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'
GAGAGAGAGCGCGCGCACTCGAGTTTCTTTTCTTTTAA 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'
GAGAGAGAGATTCTCGACTTAAATTAATTAATCCCGCCCGCC 3'], cDNA was cleaved with BamHI and XhoI. Vector: a modified pluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 83 a 104 c 57 g 44 t
ORIGIN

Query Match 64.0%; Score 16; DB 160; Length 288;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 cgtcagtgacagtcac 17
|||||
Db 188 CCTCAGTCAGTCAGTC 173

RESULT 7
W54143 328 bp mRNA EST 03-JUN-1996
LOCUS m54143.1 Soares mouse embryo NDMEL3.5 14.5 Mus musculus cDNA
DEFINITION clone IMAGE:368260 5', mRNA sequence.
ACCESSION W54143
VERSION W54143.1 GI:1355214
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 328)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Stepoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Maria M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:229692
Seq primer: mob.REGA+ET
Location/Qualifiers

FEATURES
Location/Qualifiers

source

1. .328
/organism="Mus musculus"
/strain="C57BL/6j"
/db_xref="taxon:10090"
/clone="IMAGE:368260"
/clone_1ib="Soares mouse embryo NDMEL3.5 14.5"
/sex="unknown"
/tissue-type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTCACATCTGAGTCGAGCGCGCGCGCAATTTTCTTTTCTTTTCTTTT 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Falima Bernaldo."

BASE COUNT 81 a 66 c 85 g 96 t
ORIGIN

Qy 4 tcaagtgacagtcacg 19
|||||
Db 86 TCAGTCAGTCAGTCAGC 101

RESULT 8
W34110 338 bp mRNA EST 13-MAY-1996
LOCUS m57C06.r1 Soares mouse embryo NDMEL3.5 14.5 Mus musculus cDNA
DEFINITION clone IMAGE:352618 5', mRNA sequence.
ACCESSION W34110
VERSION W34110.1 GI:1316081
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 338)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Stepoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Maria M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:224418
Putative full length read
Seq primer: mob.REGA+ET
High quality sequence stop: 324.
Location/Qualifiers
1. .338
/organism="Mus musculus"
/strain="C57BL/6j"
/db_xref="taxon:10090"

FEATURES
source

Location/Qualifiers

BASE COUNT	86 a	65 c	84 g	103 t
ORIGIN	/clone="IMAGE:352618" /clone_lib="Soares mouse embryo NBMEL3.5 14.5" /sex="unknown" /tissue_type="embryo" /dev_stage="13.5-14.5dpc total fetus" /lab_host="DH10B" /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - 0190(dt) primer [5', TGGTACCAATCGAAGGAGCGCGCGCGAATTTTTTTTTTTTTTTT 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2 1; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Ronaldo."			
Query Match	Best Local Similarity	64.0%;	Score 16;	DB 190;
Matches	16; Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;	
Db	82	TCAGTGCAGCTCAGC	97	
QY	4	tcagtgacagtcagc	19	
LOCUS	AV633196/c	345 bp	mRNA	EST
DEFINITION	AV633196 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii	15-DEC-2000		
ACCESSION	AV633196			
VERSION	AV633196.1	GI:10776516		
KEYWORDS	EST.			
SOURCE	Chlamydomonas reinhardtii.			
ORGANISM	Chlamydomonas reinhardtii.			
REFERENCE	Eukaryote; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;			
AUTHORS	Chlamydomonadaceae; Chlamydomonas.			
TITLE	1 (bases 1 to 345)			
JOURNAL	Asamizu, E., Miura, K., Kuchio, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,			
MEDLINE	Nakamura, Y., and Tabata, S.			
COMMENT	Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii			
FEATURES	DNA Res. 7 (5), 303-307 (2000)			
SOURCE	Contact: Erika Asamizu			
	The First Laboratory for Plant Gene Research			
	Kazusa DNA Research Institute			
	Yana 1532-3, Kisarazu, Chiba 292-0812, Japan			
	Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.			
	Location/Qualifiers			
	1..345			
	/organism="Chlamydomonas reinhardtii"			
	/strain="C9"			
	/db_xref="taxon:3055"			
	/clone="HC017H07_r"			
	/clone_lib="Chlamydomonas reinhardtii 5% CO2"			
	/note="Vector: pluscriptII SR-; Site_1: EcoRI; Site_2:			
	XhoI; The cDNA library was constructed from cells cultured			
	in a medium with bubbling air containing 5% carbon			
	dioxide"			
BASE COUNT	62 a	105 c	98 g	79 t
ORIGIN	1 others			
Query Match	Best Local Similarity	64.0%;	Score 16;	DB 31;
Matches	16; Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;	

RESULT	10
LOCUS	AA110497/c
DEFINITION	mle2e0j.r1 Strataegoe mouse testis (#937308) Mus musculus cDNA
ACCESSION	AA110497
VERSION	AA110497.1
KEYWORDS	GI:1662274
SOURCE	EST.
ORGANISM	house mouse. Mus musculus
REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. Ennalyola; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomi; 1 (bases 1 to 368)
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubugue,T., Schellenberg,K., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Gaisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Theising,B., Wylie,T., Lennon,G., Soares,B., Willson,R. and Waterston,R.
TITLE	The WashU-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:310452
FEATURES	Seq primer: -28m13 rev1 ET from Amersham.
SOURCE	Location/Qualifiers 1..368 /organism="Mus musculus" /strain="Inbred CD-1" /db_xref="taxon:10090" /clone="IMAGE:516604" /clone_id="Strataegoe mouse testis (#937308)" /sex="males" /tissue_type="testis" /dev_stage="10-12 week old" /lab_host="SOLR (Kanamycin resistant)" /note="Organ: testis; Vector: Bluescript SK-; Site_1: ECORI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAAATCGCCAGCAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT	75 a 115 c 106 g 72 t
ORIGIN	
Query Match	64.0%; Score 16; DB 2; Length 368;
Best Local Similarity	100.0%; Pred. No. 70;
Matches 16; Conservative	0; Mismatches 0; Indels 0; Gaps 0
OY	3 gtccagtgacagtcag 18
Db	21 GTCAGGTGCAGGTCCAG 6
RESULT	11
LOCUS	AW822106
DEFINITION	uw13c09.x1 Ren Stubbs mouse thymus Mus musculus cDNA clone IMAGE:2802352 3', mRNA sequence.
ACCESSION	AW822106
VERSION	AW822106.1
GI:	7915183

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae: Mus.
1 (bases 1 to 370)
Marras, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Rittler, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Other_ESTs: ug13c09.y1
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LINDA; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1041876
Seq primer: Primer name ambiguous
High quality sequence stop: 303.
Location/Qualifiers

FEATURES
Source
1..370
/organism="Mus musculus"
/strain="C3H"
/db_xref="taxon:10090"
/clone="IMAGE:2802352"
/clone_lib="Ren Stubbs mouse thymus"
/sex="mixed"
/dev_stage="3 weeks"
/lab_host="DH10B"
/note="Organ: thymus; Vector: pT773D-Pac; Site_1: NotI; Site_2: PacI; 1st strand cDNA was primed with an oligo(dT) primer; double-stranded cDNA was ligated using 5' linker gccgcctat and 3' linker aacgaagcttcaatt. Library is size-selected >2.5 kb and average insert size is 3.5 kb. Clones were arrayed from primary plating; non-amplified. Library constructed by X. Ren and L. Stubbs (Lawrence Livermore National Laboratory and DOE Joint Genome Institute, 7000 East Ave, L-453, Livermore, CA 94550)."

BASE COUNT
125 a 57 c 75 g 113 t

ORIGIN

Query Match 64.0%; Score 16; DB 121; Length 370;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gtcaggtgcaggtcag 18
|||||

Db 53 gtcaggtgcaggtcag 68

RESULT 12
BF444068 395 bp mRNA EST 01-DEC-2000
LOCUS 261798 MARC 2PTG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BF444068
VERSION BF444068.1 GI:11504160
KEYWORDS EST.
SOURCE Pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 395)
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W., and Keele, J.W.
Design and use of two pooled tissue normalized cDNA libraries for

JOURNAL
COMMENT Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.
PCR primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCTCCAGTCACGACG
Plate: 96 row: D column: 1
Seq primer: ATTAGGTGACACTATGAG.
Location/Qualifiers

FEATURES
Source
1..395
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2PTG"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPOR6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from testis, ovary, endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT
118 a 66 c 143 g 68 t

ORIGIN

Query Match 64.0%; Score 16; DB 148; Length 395;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 caggtgcaggtcagca 20
|||||

Db 309 CAGGTGCAGGTGACGACCA 324

RESULT 13
AQ298728 398 bp DNA GSS 15-DEC-1998
LOCUS HS_3110_B1.C12_MR_C1T Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=3110 COL=23 Row=F, DNA sequence.
ACCESSION AQ298728
VERSION AQ298728.1 GI:4015907
KEYWORDS GSS.
SOURCE Homo sapiens
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 398)
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3110 row: F column: 23
Class: BAC ends
High quality sequence stop: 398.
Location/Qualifiers

FEATURES
Source
1..398
/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone_lib="CIR Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pHELoBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT      87 a      93 c      102 g      116 t
ORIGIN

Query Match      64.0%; Score 16; DB 226; Length 398;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 ggtgcaggtcagcagc 22
        |||
        330 GGTGCAGGTGTCAGCAGC 345

RESULT 14
A1467556/c      430 bp      mRNA      EST      09-MAR-1999
LOCUS
DEFINITION      ve36909.x1 Soares_mammary_gland_NBMG Mus musculus cDNA clone
ACCESSION      A1467556
IMAGE:820288 3', mRNA sequence.
VERSION      A1467556.1 GI:4320893
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 430)
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ccgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
This clone is available royalty-free through LNL: contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:488568
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
High quality sequence stop: 406.
Location/Qualifiers
1..430
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="IMAGE:820288"
/clone_lib="Soares_mammary_gland_NBMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia
) with a modified polylinker. Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5'
TGTACCAATCTGAACTGGAGCGCGCGGAATGCTTTTCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was made by cloning products
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT      136 a      102 c      85 g      107 t
ORIGIN

Query Match      64.0%; Score 16; DB 20; Length 430;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      4 tcaagtgacagtcagc 19
        |||
        259 TCAAGTGACAGTCAGC 244

RESULT 15
AW892602
LOCUS
DEFINITION      CM3-NN0004-100300-111-f07 NN0004 Homo sapiens cDNA, mRNA sequence.
ACCESSION      AW892602
VERSION      AW892602.1 GI:8056807
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 544)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
2020263
JOURNAL
MEDLINE
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-cm3-NN0004-100
300-111-f07&t3=2000-03-10&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 543.
Location/Qualifiers
1..544
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NN0004"
/dev_stage="Adult"
/note="Organ: nervous normal; Vector: puc18; site_1: SmaI;
site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      127 a      152 c      160 g      105 t
ORIGIN

Query Match      64.0%; Score 16; DB 122; Length 544;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 gtcaggtgcagtcagc 18
        |||
        383 GTCAGGTGCAGTCAGC 398

RESULT 16
AO530462/c      553 bp      DNA      GSS      18-MAY-1999
LOCUS
DEFINITION      RPCI-11-369F11.TU RPCI-11 Homo sapiens genomic clone RPCI-11-369F11
, DNA sequence.

```

ACCESSION A0330462
 VERSION A0330462.1 GI:4842505
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 553)
 AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.
 TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
 JOURNAL Unpublished (1997)
 COMMENT Other GSSs: RPCI-11-369F11.TV
 Contact: Shaying Zhao, William Nierman, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbe@tigr.org
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
 Seq primer: SP6
 Class: BAC ends.
 FEATURES
 source Location/Qualifiers
 1..553
 /organism="Homo sapiens"
 /db_xref="GDB:7641442"
 /db_xref="taxon:9606"
 /clone="RPCI-11-369F11"
 /clone_1ib="RPCI-11"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBAC6.6, Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"
 BASE COUNT 108 a 165 c 157 g 123 t
 ORIGIN
 Query Match 64.0%; Score 16; DB 229; Length 553;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 gtcaggtgcaggtcag 18
 |||
 Db 204 gtcaggtgcaggtcag 189
 RESULT 17
 BE335032 775 bp mRNA EST 09-AUG-2000
 LOCUS 601233227F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3596935 5',
 DEFINITION mRNA sequence.
 BE335032
 BE335032.1 GI:9763677
 EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 775)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Jeffrey Green M.D.
 CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM8775 row: f column: 08
 High quality sequence stop: 225.
 FEATURES
 source Location/Qualifiers
 1..775
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:3596935"
 /clone_1ib="NCI_CGAP_Mam6"
 /sex="female, virgin"
 /tissue_type="infiltrating ductal carcinoma"
 /dev_stage="5 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-Sport6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"
 BASE COUNT 170 a 186 c 221 g 198 t
 ORIGIN
 Query Match 64.0%; Score 16; DB 136; Length 775;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 cgtcaggtgcaggtcag 17
 |||
 Db 201 cgtcaggtgcaggtcag 186
 RESULT 18
 BE194624/c standard; RNA; EST; 822 BP.
 ID BE194624
 AC BE194624;
 SV BE194624.1
 XX
 XX 05-JUL-2000 (Rel. 64, Created)
 DT 30-JUL-2000 (Rel. 64, last updated, Version 2)
 XX
 XX HVSMEH0086E11f Hordeum vulgare 5-45 DAP spike EST library HVCNDA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMEH0086E11f, mRNA sequence.
 DE
 DE EST.
 XX
 XX Hordeum vulgare (barley)
 OS Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Hordeum.
 XX
 XX [1]
 RP 1-822
 RA Wing R., Close T.J., Kleinhoof A., Wise R., Begum D., Frisch D., Yu Y.,
 RA Anderson R., Dale J., Henry D., Kernodle S., Palmer M., Rambo F.,
 RA Saski C., Schwartzbeck J., Simmons J., Choi D.W., Main D., Wood T.,
 RT "Development of a genetically and physically anchored EST resource for
 RT barley genomics";
 RL Unpublished.
 CC Contact: Wing RA
 CC Clemson University Genomics Institute
 CC Clemson University
 CC 100 Jordan Hall, Clemson, SC 29634, USA
 CC Tel: 864 656 7288
 CC Fax: 864 656 4293
 CC Email: rwing@clemson.edu
 CC Seq primer: AATTAACTCTACTAAGGG
 CC High quality sequence start: 225

KEYWORDS
CAP trapper.
Mus musculus (strain:C57BL/6J) ES cells cDNA to mRNA,
clone_lib:RIKEN full-length enriched mouse cDNA library
clone:2410198H06.

ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (sites)
Carninci, P. and Hayashizaki, Y.
High efficiency full-length cDNA cloning
Methods Enzymol. 303, 19-44 (1999)

JOURNAL
2 (sites)
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

TITLE
204933/4

JOURNAL
3 (sites)
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M.,
Kikuchi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T.,
Harada, A., Yamamoto, K., Matsumoto, H., Sakaguchi, S., Ikegami, T.,
Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Iwama, M., Ohara, E.,
Matahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,
Matsunaga, S., Okazaki, Y., Muramatsu, M., Inoue, Y. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

REFERENCE
4 (sites)
The RIKEN Genome Exploration Research Group Phase II Team and
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

JOURNAL
5 (bases 1 to 1011)

**ADACHI, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
Arikawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M.,
Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Horii, F.,
Imotani, K., Ishii, Y., Itoh, M., Iwama, M., Kato, H., Kawai, J.,
Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T.,
Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Onno, M., Okazaki, Y.,
Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H.,
Sasaki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
Tahara, Y., Tejima, Y., Toyota, T., Yamamura, T., Yasunishi, A.,
Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome-gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)**

TITLE
Further details

JOURNAL
Location/Qualifiers
1. .1011
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGI:1910427"
/db_xref="MGI:1924031"
/clone="2410198H06"
/cell_type="ES cells"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
255 a 227 c 180 g 349 t

BASE COUNT
ORIGIN

Query Match 64.0%; Score 16; DB 192; Length 1011;
Best Local Similarity 100.0%; Pred. No. 68;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 gtcagtgacgtacg 18
|||||
DB 290 gtcagtgacgtacg 275

RESULT 22
BG328084/c 1298 bp mRNA EST 27-FEB-2001
602427134F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4546497 5',
mRNA sequence.
BG328084 BG328084.1 GI:13134522
EST.
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 1298)
NIH-MGC http://mgi.nci.nih.gov/
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:
http://image.llnl.gov
Plate: LUCM1233 row: c column: 10
High quality sequence stop: 511.

FEATURES
source
1. .1298
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4546497"
/clone_lib="NIH_MGC_15"
/csize_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pOTB7; Site: 1; XhoI; Site: 2;
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT
ORIGIN
316 a 498 c 357 g 127 t

Query Match 64.0%; Score 16; DB 152; Length 1298;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tgcagtgacgtacg 16
|||||
DB 984 tgcagtgacgtacg 969

RESULT 23
AA699872/c 180 bp mRNA EST 19-DEC-1997
zj81f04.s1 Soares_fetal_liver_spleen_infls_s1 Homo sapiens cDNA
clone IMAGE:461311 3', mRNA sequence.
AA699872
ACCESSION
AA699872.1 GI:2702835
VERSION
AA699872.1
KEYWORDS
EST.
SOURCE
ORGANISM
Homo sapiens

REFERENCE 1 (bases 1 to 180)
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
 TITLE Washington-NCI human EST Project
 JOURNAL Unpublished (1997)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (Info@image.llnl.gov) for further information.
 Seq primer: -40m13 fwd. ET from Amerisham
 High quality sequence stop: 144.
 Location/Qualifiers

FEATURES

source

```
1..180
/organism="Homo sapiens"
/db_xref="GDB:3752334"
/db_xref="taxon:9606"
/clone="IMAGE:461311"
/clone_lib="Soares_fetal_liver_spleen_LNLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Pac I; Site 2: Eco RI; This is a subtracted version of the original Soares fetal liver spleen LNLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGCAAGATTATTAAGATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
```

BASE COUNT

47 a 57 c 36 g 40 t

Query Match 60.0%; Score 15; DB 10; Length 180;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 tcagtgacgtcagc 19
 |||||
 Db 138 CAGTGCAGTGCAGC 124

RESULT 24

H22139 264 bp mRNA EST 06-JUL-1995
 LOCUS H22139/c
 DEFINITION Y138a03.s1 Soares breast 3nbhst Homo sapiens cDNA clone
 IMAGE:160492 3', mRNA sequence.
 ACCESSION H22139
 VERSION H22139.1 GI:890834
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE

1 (bases 1 to 264)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
 TITLE The Washu-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK

Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 685
 High quality sequence stops: 241
 Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the IMAGE Consortium (Info@image.llnl.gov) for further information.
 Insert length: 685 Std Error: 0.00
 Seq primer: Promega -2im13
 High quality sequence stop: 241.
 Location/Qualifiers

FEATURES

source

```
1..264
/organism="Homo sapiens"
/db_xref="GDB:574535"
/db_xref="taxon:9606"
/clone="IMAGE:160492"
/clone_lib="Soares breast 3nbhst"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: breast; Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAGTGGAGCGCCGCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 20. Library constructed by Bento Soares and M.Fatima Bonaldo."
```

BASE COUNT

87 a 47 c 50 g 73 t 7 others

ORIGIN

Query Match 60.0%; Score 15; DB 157; Length 264;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcagtgacgtcagc 18
 |||||
 Db 215 TCAGTGCAGTGCAGC 201

RESULT 25

R50061 279 bp mRNA EST 18-MAY-1995
 LOCUS R50061/c
 DEFINITION yf59c10.s1 Soares breast 2nbhst Homo sapiens cDNA clone
 IMAGE:153042 3', mRNA sequence.
 ACCESSION R50061
 VERSION R50061.1 GI:811963
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE

1 (bases 1 to 279)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
 TITLE The Washu-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK

TITLE

JOURNAL

Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 681
 High quality sequence stops: 257 Source: IMAGE Consortium, LNL

/lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site: 1;
 EcoRI; Site: 2; EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI Methylase. Size
 selected DNA was cloned into the pBACe3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."
 BASE COUNT 61 a 74 c 78 g 80 t
 ORIGIN

Query Match 60.0%; Score 15; DB 251; Length 293;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gtcaggtgcaggtca 17
 |||
 Db 55 gtcaggtgcaggtca 41

RESULT 28
 BB175346 300 bp mRNA EST 29-JUN-2000
 LOCUS BB175346 RIKEN full-length enriched, adult male hypothalamus Mus
 DEFINITION musculus cDNA clone A230056E21 3', mRNA sequence.
 ACCESSION BB175346
 VERSION BB175346.1 GI:8834429
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 300)
 AUTHORS Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci
 Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
 Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
 Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
 Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,
 Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
 Ono, T., Owa, C., Salto, H., Sakai, C., Sato, K., Shibata, K., Shibata,
 Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,
 Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomioka, N., Toya,
 T., Tsunoda, Y., Watanabe, S., Yamamoto, T., Yamamoto, T., Yamamoto, T.,
 Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino,
 M., Muramatsu, M., and Hayashizaki, Y.
 RIKEN Mouse ESTs (Kono, H., et al.)
 Unpublished (2000)
 CONTACT: Yoshihide Hayashizaki
 Genome Exploration Research Group, Life Science Tsukuba Center,
 The Institute of Physical and Chemical Research (RIKEN), Genomic
 Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: +81-298-36-9013
 Fax: +81-298-36-9098
 Email: genome-res@rtc.riken.go.jp,
 URL: http://genome.rtc.riken.go.jp/
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki,
 N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 Thermostabilization and thermostabilization of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh, M., Kikunishi, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
 Y., and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P., and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (http://genome.rtc.riken.go.jp) for
 further details.

FEATURES

source

Location/Qualifiers
 1..300
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="A230056E21"
 /clone_id="RIKEN full-length enriched, adult male
 hypothalamus"
 /sex="male"
 /tissue_type="hypothalamus"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Site: 1; Sali; Site: 2; BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5',
 GAGAGAGAGAGATCCACAGAGCTCTTTTCTTTTCTTTTCTTTT 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. cDNA went through one round of normalization
 to Rot = 20.0 and subtraction to Rot = 459.0. Second
 strand cDNA was prepared with the primer adapter of
 sequence [5' GAGAGAGAGATCTCGATTAAATTAATATATCCCCCCCC
 3']. cDNA was cleaved with XhoI and BamHI. Vector: a
 modified pBluescript KS(+) after bulk excision from Lambda
 FLC I."

BASE COUNT

ORIGIN

Query Match 60.0%; Score 15; DB 127; Length 300;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcaggtgcaggtcag 18
 |||
 Db 72 tcaggtgcaggtcag 58

RESULT 29
 BG100182 306 bp mRNA EST 29-JAN-2001
 LOCUS BG100182 uc94904.y1 McCarrey Eddy type B spermatogonia Mus musculus cDNA
 DEFINITION clone IMAGE:3656262 5', mRNA sequence.
 ACCESSION BG100182
 VERSION BG100182.1 GI:12595499
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 306)
 AUTHORS Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
 Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person,
 B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schunk, R., Ritter,
 E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
 Waterston, R., and Willson, R.
 The WashU-NCI Mouse EST Project 1999
 Unpublished (1999)
 CONTACT: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

TITLE
 JOURNAL
 COMMENT

Email: mouseest@wustl.edu
 This clone is available royalty-free through LNC; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:141856
 Seq primer: Primer name ambiguous.
 Location/Qualifiers
 1..306
 source

```

/organism="Mus musculus"
/strain="CD-1"
/db_xref="taxon:10090"
/clone="IMAGE:3656262"
/clone_1ib="McCarrey Eddy type B spermatogonia"
/sex="male"
/tissue_type="type B spermatogonia, pooled from multiple mice"
/dev_stage="8 day"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: testis; Vector: pBluescript SK+ (Stratagene); Site_1: XhoI; Site_2: EcoRI; CDNA oligo dt-primed (5'-(GA)10-ACGAGTCGAGTTTGT-3') and directionally cloned using 5' linkers 5'-AATCGGACAGC-3' and 5'-CTCGGCGCCG-3'. Size selection of >400bp material gives average insert size ranging from 1-2 kb. Library was mass excluded (from lambda-UniZAP-XR) and resulting single-stranded phagemids were prepped and transformed into DH10B. Library contains 96% recombinants. References: J. Androl. 20:635-639 and Gene 25:263-269. Library constructed and donated by J. McCarrey, Ph.D. (Southwest Foundation for Biomedical Research, Dept. of Genetics); excision done by E.M. Eddy, Ph.D. (National Institutes of Health, National Institute of Environmental Health Sciences). Original lambda-based library is available through ATCC, catalog #63417."
BASE COUNT      62 a      82 c      93 g      69 t
ORIGIN

```

```

Query Match      60.0%; Score 15; DB 173; Length 306;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3      gtcaggtcaggtca 17
        |||||
Db      95      gtacggtcaggtca 81

```

```

RESULT 30
LOCUS      A1138808      382 bp      mRNA      EST      28-Oct-1998
DEFINITION g989602.x1 Soares_testis_NHT Homo sapiens CDNA clone IMAGE:1737554
ACCESSION  A1138808
VERSION     A1138808
KEYWORDS    A1138808.1 GI:3644780
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 382)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: c9apbs-rt@mail.nih.gov
            CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
            CDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LINL at:
            www-bio.lnl.gov/bbrp/image/image.html
            Insert length: 593 Std Error: 0.00
            Seq primer: -40m13 fwd. RT from Amersham
            High quality sequence stop: 362.
            Location/Qualifiers
            1..382
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:1737554"

```

```

/clone_1ib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand CDNA
was prepared from mRNA obtained from Clontech laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5'
TGTTACCATCTGAAGTGGAGCGGCCGACATTTTGT-3'].
Double-stranded CDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      103 a      84 c      75 g      120 t
ORIGIN

```

```

Query Match      60.0%; Score 15; DB 16; Length 382;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      6      aggtgcaggtcagca 20
        |||||
Db      311      AGGTGACAGTCACCA 325

```

```

RESULT 31
AW522699/c
LOCUS      AW522699      385 bp      mRNA      EST      06-MAR-2000
DEFINITION UI-R-B00-ahk-c-01-0-UI.s1 UI-R-B00 Rattus norvegicus CDNA clone
ACCESSION  AW522699
VERSION     AW522699
KEYWORDS    AW522699.1 GI:7165124
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 385)
AUTHORS    Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: mscares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand CDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the CDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized medulla library CDNA library Preparation: M.B. Soares
            lab Clone distribution: clones will be available through Research
            Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
            Location/Qualifiers
            1..385
            /organism="Rattus norvegicus"
            /strain="Sprague-Dawley"
            /db_xref="taxon:10116"
            /clone="UI-R-B00-ahk-c-01-0-UI"
            /clone_1ib="UI-R-B00"
            /dev_stage="adult"
            /lab_host="DH10B (Life Technologies)"
            /notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
            polylinker; Site_1: Not I; Site_2: Eco RI; The library

```

(UI-R-B00) is a subtracted library derived from a mixture of the following tissues: thalamus, cerebellum, hypothalamus, medulla, pons, midbrain, cerebral cortex, corpus striatum and hippocampus. For a detailed description of the library from which this clone was derived, please visit our web site at tatest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)

6:791-806, 1996)
TAG: LIB-UI-R-B00
TAG: TISSUE=medulla
TAG: SEQ=CAACCG
BASE COUNT 87 a 104 c 98 g 96 t
ORIGIN

Query Match 60.0%; Score 15; DB 117; Length 385;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 gtcaggtcagtcacg 22
|||||
Db 336 CTCAGGTGACGACG 322

RESULT 32
AA062258 396 bp mRNA EST 10-JUL-2000
LOCUS 69428 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION
ACCESSION AA062258
VERSION AA062258.1 GI:7056146
KEYWORDS EST.
SOURCE
ORGANISM

Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 396)
Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
M.W. and Keefe,J.W.
Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle

JOURNAL
COMMENT Unpublished (2000)

USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross-match with the -munscore 20
and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCGCCAGTCACGACG
Plate: 45 row: 1 column: 7
Seq primer: ATTTAGGTGACACTATAG.

FEATURES
source

1..396
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 4BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 20 and day 40
embryos."
BASE COUNT 70 a 134 c 127 g 65 t
ORIGIN

Query Match 60.0%; Score 15; DB 116; Length 396;

Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 caggtcagtcacg 19
|||||
Db 221 CAGGTGACGTCACG 207

RESULT 33
AA062258 407 bp mRNA EST 03-FEB-1997
LOCUS m155908.t1 Stratagene mouse testis (#937308) Mus musculus cDNA
DEFINITION
ACCESSION AA062258
VERSION AA062258.1 GI:1556057
KEYWORDS EST.
SOURCE
ORGANISM

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognath; Muridae; Murinae; Mus.
1 (bases 1 to 407)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMNI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMNI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
MGI:309814

Seq primer: -26m13 rev1 ET from Amersham
High quality sequence stop: 387.

FEATURES
source

1..407
/organism="Mus musculus"
/strain="Inbred CD-1"
/db_xref="taxon:10090"
/clone_image="515966"
/clone_lib="Stratagene mouse testis (#937308)"
/sex="males"
/tissue_type="testis"
/dev_stage="10-12 week old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: testis; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
-5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTGTGTTTTTTTTTTT 3'."
BASE COUNT 88 a 117 c 123 g 79 t
ORIGIN

Query Match 60.0%; Score 15; DB 1; Length 407;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gtcaggtcagtcacg 17
|||||
Db 15 CTCAGGTGACGTCACG 1

RESULT 34
BF599674 407 bp mRNA EST 13-DEC-2000
LOCUS BF599674/C
DEFINITION 263546 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.

ACCESSION BF599674
 VERSION BF599674.1 GI:11696393
 KEYWORDS EST.
 SOURCE COW.
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
 1 (bases 1 to 407)
 Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid,W.M. and Keele,J.W.
 Design and use of four pooled tissue normalized cDNA libraries for EST discovery in cattle
 Unpublished (2000)
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -m1nscore 18 and -mismatch 12 options.
 PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCACAGTCACGAG
 Plate: 35 row: E column: 3
 Seq primer: ATTGAGTGACACTATAG.
 Location/Qualifiers
 1..407
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 3BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle."
 BASE COUNT 69 a 139 c 135 g 64 t
 ORIGIN
 Query Match 60.0%; Score 15; DB 151; Length 407;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 caggtgcaggtcagc 19
 |||||||||||||
 Db 93 CAGGTGCAGGTACG 79
 RESULT 35
 A0085001 408 bp DNA GSS 26-AUG-1998
 LOCUS HS.2269_A2.B10_MR CIT Approved Human Genomic Sperm Library D Homo
 DEFINITION sapiens genomic clone Plate=2269 Col=20 Row=C, DNA sequence.
 ACCESSION A0085001
 VERSION A0085001.1 GI:3454218
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 408)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 99380589
 TITLE
 JOURNAL
 MEDLINE

COMMENT
 Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 2269 row: C column: 20
 Class: BAC ends
 High quality sequence stop: 408.
 Location/Qualifiers
 1..408
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_plate="Plate=2269 Col=20 Row=C"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coll DH10B"
 BASE COUNT 104 a 111 c 78 g 115 t
 ORIGIN
 Query Match 60.0%; Score 15; DB 224; Length 408;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 caggtgcaggtcagc 19
 |||||||||||||
 Db 396 CAGGTGCAGGTACG 382
 RESULT 36
 BF653430 409 bp mRNA EST 20-DEC-2000
 LOCUS 272257 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
 DEFINITION BF653430
 ACCESSION BF653430
 VERSION BF653430.1 GI:11918562
 KEYWORDS EST.
 SOURCE cow.
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
 1 (bases 1 to 409)
 Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid,W.M. and Keele,J.W.
 Design and use of four pooled tissue normalized cDNA libraries for EST discovery in cattle
 Unpublished (2000)
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -m1nscore 18 and -mismatch 12 options.
 PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCACAGTCACGAG
 Plate: 69 row: P column: 10
 Seq primer: ATTGAGTGACACTATAG.
 Location/Qualifiers
 1..409
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 3BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"


```

RESULT 39
LOCUS BE604067 457 bp mRNA EST 13-DEC-2000
DEFINITION B604067 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BF604067
VERSION BF604067.1 GI:11702307
KEYWORDS EST.
SOURCE cow
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 457)
Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
,W.M. and Keeler,J.W.
Bennett,G.A., Fahrrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
,W.M. and Keeler,J.W.
Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
Unpublished (2000)
JOURNAL Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR primers
FORWARD: AGGAACAGCATATGACCAT
BACKWARD: GTTTCACAGTCACGACG
Plate: 50 row: 1 column: 16
Seq primer: ATTAGTGACACTATAG.
FEATURES
source 1. 457
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site.1: XbaI; Site.2: XhoI;
library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."
BASE COUNT 82 a 152 c 142 g 80 t 1 others
ORIGIN
Query Match 60.0%; Score 15; DB 151; Length 457;
Best Local Similarly 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 caggtgcagtcagc 19
|||||
Db 215 CAGGTGCAGTCAGC 201

```

```

TITLE and Fraser,C.M.
JOURNAL Mouse BAC End Sequences from Library RPCI-23
COMMENT Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.bufileo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.bufileo.edu/orderingframe.htm)
or from Resea ch Genetics (http://www.tigr.org/tldb/bac-ends/mouse/bac\_end\_intro.html)
plate: 273 row: P column: 15
Seq primer: T7
Class: BAC ends.
FEATURES
source 1. 477
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-273P15"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site.1:
EcoRI; Site.2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 101 a 107 c 125 g 144 t
ORIGIN
Query Match 60.0%; Score 15; DB 235; Length 477;
Best Local Similarly 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 gtcaggtgcagtcac 17
|||||
Db 68 GTCAGGTGCAGTCAC 54

```

Single pass sequencing. Bases called and alt-trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCCAGTCAGCAGC
 Plate: 68 row: P column: 2
 Seq primer: ATTGAGTCACCTATAG.

FEATURES

Location/Qualifiers
 1..492

/organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 2BOY"
 /tissue_type="pooled"
 /lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from testis, thymus,
 semitendinosus muscle, longissimus muscle, pancreas,
 adrenal, and endometrium."

BASE COUNT 96 a 161 c 149 g 85 t 1 others
 ORIGIN

Query Match 60.0%; Score 15; DB 140; Length 492;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 caggtgcaggtcag 19
 |||
 DB 263 CAGGTGCAGGTGAC 249

RESULT 42

AZ261270 495 bp DNA GSS 26-JUL-2000
 LOCUS AZ261270/c
 DEFINITION RPI-23-44018.TV RPI-23 Mus musculus genomic clone RPI-23-44018,
 DNA sequence.

ACCESSION AZ261270
 VERSION AZ261270.1 GI:9469499
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 495)
 Authors Zhao, S., Nieman, W., Feldblyum, T., Malek, J., Shatsman, S., Akiret,
 B., Levins, M., Mogan, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.,
 and Fraser, C.M.

TITLE Mouse BAC End Sequences from Library RPI-23
 JOURNAL Unpublished (1999)
 COMMENT Other GSSs: RPI-23-44018.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@igr.org

Clones are derived from the mouse BAC library RPI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieterdejong.med.bufileo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.bufileo.edu/orderingframe.htm)
 or from Resea ch Genetics (info@resgen.com). BAC end page:
 http://www.tigr.org/tdb/bac/ends/mouse/bac_end_intro.html
 Plate: 440 row: I column: 8
 Seq primer: T7

FEATURES

Class: BAC ends.
 Location/Qualifiers

1..495
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPI-23-44018"

/clone_lib="RPI-23"
 /sex="Female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1:
 EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI methylase. Size
 selected DNA was cloned into the pBAC3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 106 a 110 c 132 g 147 t
 ORIGIN

Query Match 60.0%; Score 15; DB 240; Length 495;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 gtcaggtgcaggtc 17
 |||
 DB 66 GTCAGGTGCAGGTCA 52

RESULT 43
 BF652819/c 496 bp mRNA EST 20-DEC-2000
 LOCUS BF652819
 DEFINITION 276504 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION BF652819
 VERSION BF652819.1 GI:11917951
 KEYWORDS EST.
 SOURCE COW.

ORGANISM

Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 496)

REFERENCE

Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grose, W.M.,
 Bennett, G.A., Fahrenkrug, S.C., Freking, B.A., Kohrer, G.A., Laegreid,
 W.W., and Keeler, J.W.

Design and use of four pooled tissue normalized cDNA libraries for
 EST discovery in cattle
 Unpublished (2000)
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@maill.marc.usda.gov

Single pass sequencing. Bases called and alt-trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -mismatch 12 options.
 PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCCAGTCAGCAGC
 Plate: 67 row: C column: 19
 Seq primer: ATTGAGTCACCTATAG.

FEATURES

source

Location/Qualifiers

1..496
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 3BOY"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from marrow, alveolar
 macrophage, ovary, fetal semitendinosus muscle, and fetal
 longissimus muscle."

BASE COUNT 100 a 164 c 148 g 84 t
 ORIGIN

Query Match 60.0%; Score 15; DB 151; Length 496;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 caggtgcaggtcagc 19
 |||
 DB 290 CAGGTGCAGGTACG 276

RESULT 44

BF567735

LOCUS BF567735 525 bp mRNA EST 12-DEC-2000
 DEFINITION UT-R-B00-ahk-c-01-0-UT.1 UT-R-B00 Rattus norvegicus cDNA clone

ACCESSION BF567735

VERSION BF567735.1 GI:11677465

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 525)

AUTHORS Bonaldo, M.F., Lennon, G., and Soares, M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

JOURNAL discovery

MEDLINE Genome Res. 6 (9), 791-806 (1996)

COMMENT Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

cDNA Library Preparation: M.B. Soares Lab Clone distribution:

clones will be available through Research Genetics (www.resgen.com)

This clone is also available through the I.M.A.G.E. Consortium at

LIML (info@liml.gov). IMAGE ID= 1796684

Seq primer: M13 Forward.

Location/Qualifiers

1..525

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UT-R-B00-ahk-c-01-0-UT"

/clone_lib="UT-R-B00"

/dev_stage="adult"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker. Site_1: Not I; Site_2: Eco RI; The library

(UT-R-B00) is a subtracted library derived from a mixture

of the following tissues: thalamus, cerebellum,

hypothalamus, medulla, pons, midbrain, cerebral cortex,

corpus striatum and hippocampus. For a detailed

description of the library from which this clone was

derived, please visit our web site at

ratseq.eng.uiowa.edu. The subtraction has been previously

described in (Bonaldo, Lennon and Soares, Genome Research

6:791-806, 1996)"

BASE COUNT 131 a 143 c 138 g 112 t 1 others

ORIGIN

Query Match 60.0%; Score 15; DB 150; Length 525;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 gtgcaggtcagcagc 22
 |||
 DB 491 GTGCAGGTACGACG 505

RESULT 45
 AM290780/c

LOCUS AM290780 529 bp mRNA EST 16-JAN-2000
 DEFINITION NXNV046H01f Nsf Xylem Normal wood Vertical Pinus taeda cDNA clone

ACCESSION NXNV046H01 5', mRNA sequence.

VERSION AM290780

KEYWORDS AM290780.1 GI:6697416

SOURCE EST.

ORGANISM 10biolly pine.

Pinus taeda

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.

1 (bases 1 to 529)

REFERENCE Sederoff, R.

Molecular Basis of Wood Formation in the Pine Megagenome

Unpublished (2000)

Contact: Johnson, Arthur

North Carolina State University

Tel: 919 515 7800

Fax: 919 515 7801

Email: ajohnson@unity.ncsu.edu

Seq primer: T3.

Location/Qualifiers

1..529

/organism="Pinus taeda"

/db_xref="taxon:3352"

/clone="NXNV046H01"

/clone_lib="Nsf Xylem Normal wood Vertical"

/note="Vector: Bluescript SK; Site_1: Eco RI; The

sequences contain a 'cDNA adapter' between the EcoRI site

and the start of the EST. The adapter sequence is

'AATTCGCGACGAG'."

BASE COUNT 157 a 89 c 129 g 133 t 21 others

ORIGIN

Query Match 60.0%; Score 15; DB 114; Length 529;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 tgcaggtcagcagc 23
 |||
 DB 65 TGCGGTACGACGCT 51

Search completed: October 9, 2001, 15:15:40
 Job time: 13662 sec

Wed Oct 10 07:45:38 2001

us-09-396-196f-10.oli.rst

Page 23

09/396.196

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:43:40 ; Search time 1670.83 Seconds
(without alignments)
231.438 Million cell updates/sec

Title: US-09-396-196f-2
Perfect score: 25
Sequence: 1 atgtgcagtcacagatatatt 25

Scoring table: OLIGO-NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 773874588 residues

Word size : 9

Total number of hits satisfying chosen parameters: 208678

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba1:*

2: gb_ba2:*

3: gb_ba3:*

4: gb_in1:*

5: gb_in2:*

6: gb_in3:*

7: gb_om:*

8: gb_ov:*

9: gb_pat1:*

10: gb_pat2:*

11: gb_ph:*

12: gb_pl1:*

13: gb_pl2:*

14: gb_pl3:*

15: gb_pl4:*

16: em_ba1:*

17: em_ba2:*

18: em_fun:*

19: em_htgo_hum:*

20: em_htgo_inv:*

21: em_htgo_rod:*

22: em_htg_hum1:*

23: em_htg_hum2:*

24: em_htg_hum3:*

25: em_htg_hum4:*

26: em_htg_hum5:*

27: em_htg_hum6:*

28: em_htg_hum7:*

29: em_htg_hum8:*

30: em_htg_inv1:*

31: em_htg_inv2:*

32: em_htg_other:*

33: em_htg_rod:*

34: em_hum1:*

35: em_hum2:*

36: em_hum3:*

37: em_hum4:*

38: em_hum5:*

39: em_hum6:*

40: em_hum7:*

41: em_in:*

42: em_om:*

43: em_or:*

44: em_ov:*

45: em_pat:*

46: em_ph:*

47: em_pl:*

48: em_ro:*

49: em_sts:*

50: em_sy:*

51: em_un:*

52: em_vl:*

53: gb_sts1:*

54: gb_sts2:*

55: gb_sts3:*

56: gb_sy:*

57: gb_un:*

58: gb_vl:*

59: gb_vl2:*

60: gb_htg1:*

61: gb_htg2:*

62: gb_htg3:*

63: gb_htg4:*

64: gb_htg5:*

65: gb_htg6:*

66: gb_htg7:*

67: gb_htg8:*

68: gb_htg9:*

69: gb_htg10:*

70: gb_htg11:*

71: gb_htg12:*

72: gb_htg13:*

73: gb_htg14:*

74: gb_htg15:*

75: gb_htg16:*

76: gb_htg17:*

77: gb_htg18:*

78: gb_htg19:*

79: gb_htg20:*

80: gb_htg21:*

81: gb_htg22:*

82: gb_htg23:*

83: gb_htg24:*

84: gb_htg25:*

85: gb_pr1:*

86: gb_pr2:*

87: gb_pr3:*

88: gb_pr4:*

89: gb_pr5:*

90: gb_pr6:*

91: gb_pr7:*

92: gb_pr8:*

93: gb_pr9:*

94: gb_r01:*

95: gb_r02:*

96: gb_in4:*

97: gb_pr10:*

98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	25	100.0	1041 9 AR029499	AR029499 Sequence
2	25	100.0	1041 9 AR034916	AR034916 Sequence
3	25	100.0	1084 9 A11350	A11350 RibB gene o
4	25	100.0	1121 10 E00893	E00893 Genomic DNA
5	25	100.0	5793 2 ECOBIO	J04423 E.coli 7,8-
6	25	100.0	5872 9 A38246	A38246 Sequence 1
7	25	100.0	5872 9 A38251	A38251 Sequence 6
8	25	100.0	5872 9 A93674	A93674 Sequence 1

```

9      25 100.0 5872 9 A93679
10     25 100.0 5872 9 AR101809
11     25 100.0 5872 9 AR101810
12     25 100.0 11022 1 AE000180
13     25 100.0 13501 1 AE005258
14     25 100.0 297816 2 AE002553
15     24 96.0 5526 2 AE250776
16     17 68.0 25256 91 AP000954
17     17 68.0 32767 77 AC087305
18     17 68.0 60072 62 AC011937
19     17 68.0 86757 84 HS5418
20     17 68.0 115551 90 AL353582
21     17 68.0 147706 12 AC083945
22     17 68.0 155172 66 AC021110
23     17 68.0 174815 82 AP001390
24     17 68.0 228516 85 AC002523
25     17 68.0 340000 91 AP001690
26     16 64.0 605 93 HSAPAF20
27     16 64.0 1829 54 G07181
28     16 64.0 4602 91 D13644
29     16 64.0 5469 3 L1A012388
30     16 64.0 43741 90 AL161800
31     16 64.0 68894 75 AC074073
32     16 64.0 105937 12 AC012679
33     16 64.0 110525 85 AB020878
34     16 64.0 11312 87 AC011248
35     16 64.0 118636 82 AL590387
36     16 64.0 129010 90 AL159168
37     16 64.0 139218 65 AC019289
38     16 64.0 152452 77 AC090105
39     16 64.0 153221 73 AC067998
40     16 64.0 155211 65 AC019258
41     16 64.0 156184 71 AC041018
42     16 64.0 158408 90 AL391821
43     16 64.0 167110 67 AC022621
44     16 64.0 167176 68 AC024560
45     16 64.0 167803 61 AC010816

```

ALIGNMENTS

```

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
source location/Qualifiers
1..1041
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 6e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 atgtgcgaagtcacagaattatt 25
|||||
Db 24 ATGTGCGCAAGTCACAGATTATT 48

```

RESULT 2

```

AR034916
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
source location/Qualifiers
1..1041
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 6e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 atgtgcgaagtcacagaattatt 25
|||||
Db 24 ATGTGCGCAAGTCACAGATTATT 48

```

```

RESULT 3
LOCUS A11530 1084 bp DNA
DEFINITION BioB gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 1084)
AUTHORS Patent: GB 2216530-A 16 11-OCT-1989;
JOURNAL Location/Qualifiers
FEATURES
source 1..1084
gene /organism="Escherichia coli"
24..1064
24..1064
CDS /gene="BioB"
24..1064
/gene="BioB"
/codon_start=1
/transl_table=1
/protein_id="CA00965.1"
/db_xref="GI:490219"
/translacion="MAHRPMTLSQVLEFEKPLDILFEAQQVHROHPPROYST
LSTKTGACPEDCKYCPOSSRYRTGLEAEELMEVYLEEAKAKAKAGSTRFGAM
KNPHERMPYLEOVGVKAMGAEACHTLTLSQKORLANAGLDYINHLDTSPF
YGNITTRTYOERLDLTLEKVRDAGIKVSGGIVGLGTYVDRAGLLIQLANLTPES
VPINMLVKVGTPLADNDVDADFDPIFTIIVARIMPTSYRLSAGREQNNBQTOAMC
FMAGANSIFYGCKILTTTPNEDEKDLQLFKRLGILNPQFVLAGDNEDQQRLEQALMT
PDTDEYYNAAL"

```

```

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 1084;
Best Local Similarity 100.0%; Pred. No. 5.9e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 atgtgcgaagtcacagaattatt 25
|||||

```

DB 47 ATTGTGCAAGTCACAGATTATT 71

RESULT 4

LOCUS E00893 1121 bp DNA PAT 29-SEP-1997

DEFINITION Genomic DNA encoding biotin Synthetase.

ACCESSION E00893

VERSION E00893.1 GI:2169154

KEYWORDS JP 1986149091-A/1.

SOURCE Escherichia coli.

ORGANISM Escherichia coli

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

Escherichia.

REFERENCE 1 (bases 1 to 1121)

AUTHORS Hirono, Y., Kojima, T. and Kimura, H.

TITLE DUPLEX DNA TO CODE BIOTIN SYNTHASE. BACTERIUM CONTAINING SAME AND PRODUCTION OF BIOTIN

JOURNAL Patent: JP 1986149091-A 1 07-JUL-1986;

COMMENT NIPPON SODA CO LTD

OS Escherichia coli

PN JP 1986149091-A/1

PD 07-JUL-1986

PF 24-DEC-1984 JP 1984272605

PI HIRONO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC

C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12P13/19),(C12P13/18,PC C12R1:19);

CC strandedness: Double;

CC topology: linear;

CC hypothetical: No;

CC anti-sense: No;

CC *source: strain=Escherichia coli NS101;

CC feature is identified by experimental;

FF Key location/Qualifiers

FF FT CDS 42..1079

FEATURES

source Location/Qualifiers

1..1121

/organism="Escherichia coli"

/db_xref="taxon:562"

BASE COUNT 289 a 296 c 325 g 211 t

ORIGIN

Query Match 100.0%; Score 25; DB 10; Length 1121;

Best Local Similarity 100.0%; Pred. No. 5.9e-05;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 attgtcgcaagtcacagaattatt 25

|||||

Db 65 ATTGTGCAAGTCACAGATTATT 89

RESULT 5

ECOBIO 5793 bp DNA BCF 28-FEB-1994

LOCUS

DEFINITION E.coli 7,8-diamino-pelargonic acid (bica), biotin synthetase (bicaB), 7-keto-8-amino-pelargonic acid synthetase (bicaF), bica protein, and dehydrobiotin synthetase (bicaD), complete cds.

ACCESSION J04423

VERSION J04423.1 GI:145422

KEYWORDS 7,8-diamino-pelargonic acid aminotransferase; 7-keto-8-amino-pelargonic acid synthetase; bica gene; bicaF gene; bicaD gene; bicaF gene; bicaF gene; biotin synthetase; dehydrobiotin synthetase.

SOURCE Escherichia coli (strain K-12) DNA.

ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 5793)

AUTHORS Otsuka, A.J., Buonocristiani, M.R., Howard, P.K., Flamm, J. and Johnson, O.

TITLE The Escherichia coli biotin biosynthetic enzyme sequences

JOURNAL J. Biol. Chem. 263, 19577-19585 (1988)

MEDLINE 89066784

COMMENT Draft entry and computer-readable sequence [1] kindly submitted by A.Otsuka, 09-NOV-1988.

FEATURES

source Location/Qualifiers

1..5793

/organism="Escherichia coli"

/strain="K-12"

/db_xref="taxon:562"

/complement(98..574)

/note="ORF 1"

/codon_start=1

/transl_table=1

/protein_id="AA23513.1"

/db_xref="GI:455168"

/translation="MKLISNDLRDGRKLPHRHVFNKGKDGDNISPHLANDVPAGTKSFVATCYDPDAPGTSGMMHWVYVNLPAIDTRVLPDGFSGSLVAMPDGVLTQRTDFGKTGYDGAAPKPGETHRHYLFYVHALDIERIDVDEGASAAVGFVHFHSLASISITAMFS"

/complement(633..1925)

/gene="bica"

/complement(633..1925)

/gene="bica"

/codon_start=1

/transl_table=1

/product="7,8-diamino-pelargonic acid aminotransferase"

/protein_id="AA23514.1"

/db_xref="GI:457106"

/translation="MTTDLAFDGRHWHPTSMTPSLPYVPVSAEGCELLSDGRIVDGSSWMAAIHGYNHPOLNNAKMSQIDAMSHVFGGITHPAIELCRLVANSGNALCEGYLADSGSVAEVAEMKMALOYQWAKGARGRFLTFRNGYHGDTEGASVCDPN SMHSLMKYGLPENLEFAPAPOSMDQEMERMGVARGMAHRIHIAVILIEPVQGA GMRVYHPBYKRIKRIKIDCGILLIADEIATGGRGKLPACAEHAIETAPILICIGA LTGGTMTSATITTRVAVETISNAGCGFMHGPMPGMPGLCAANASIAITLESQDMO QOVADIEVQLEQLAPARDAEIVADRVYLVADAVNAVADETFFCO"

/protein_id="P0010R"LTAAVNAVADETFFCO"

2012..3052

/gene="bicaB"

2012..3052

/gene="bicaB"

/codon_start=1

/transl_table=1

/product="biotin synthetase"

/protein_id="AAA23515.1"

/db_xref="GI:145425"

/translation="NAHPRRWTLQVTELEFEKPLDILFEAQOVHROHEDDPPOVSTLUSIKTGACPECKTCIPOTSRIKTGIEAEERLMEVOYLESARAKAAGSTRPCGAAY KNPHERMDPYLEDMOVGVKAMGLACMTLGTLSQAOURLANAGDITYNNHLDTSPEF YGNILITTRVYQERLDTLEKVRDAGIKVSGGIVGIGETVKDAGILLDANLPTPES VPIINMLVKVKGPTLADNDVDVADFIRTIJAVVARIIMPSTSVRLSAGRDMQEQIAMC PMAGANSIFPGCKLITTPNPEEDKDLQFLFKLGLNPDQTAVALAGDNEOORLEQALMT PTDDEYNNAL"

3049..4203

/gene="bicaF"

3049..4203

/gene="bicaF"

/codon_start=1

/transl_table=1

/product="7-keto-8-amino pelargonic acid synthetase"

/protein_id="AA23516.1"

/db_xref="GI:145426"

/translation="MSWOKRINALDARRADALRRRYVQAQAGRWLVADDRQYLNFSMDYGLSHHPQIIRAMQGAEOFGISGSGVSGVSVQALAEELAEVILGYSRALIFITSPANNOAVITAMMAKEDRIADRLSHASLEAASLSPQLRFRHNDVTHLAR LASPPGQOMVYTEGVPFSGMDGSAPLAEIOOVTOOHGMLWVDAHGGVIGEOGRSCLQKVRKPELVVTTGKGFVSGAAVILCSSYVADYLLQFARHLIYSTMPPAOAL RASIAVIRSDGEGARREKLAALITFRAGVDLPFLADSCSAIOPLIYGDNSRALQL AKRLQOOGVWVTAIRPPTVPAGTARLRLTAAHEMODIDRLLEYLHNG"

4190..4945

/gene="bicaC"

4190..4945

/gene="bicaC"

/gene="bicaC"

```

/codon_start=1
/transl_table=11
/protein_id="AA23517.1"
/db_xref="GI:145427"
/translation="MATVAKQAIAAFGRAAAHYEHOADLOROSADALLAMLPORRYT
HYLDAGCGPMMSRHMRERHQAQVATLDSPLMIVQARQKADADHYIAGDIESLPLATA
TFDLAMSNLAVOWCGLNSTALRELRYRVAPKGVYAFPTTVGSLPERHQMQAVDERP
HANRFLPPDEITEOSINQVYOHHTOPTITLWPDDELASMRSLKIGATHLHGRPRIL
TRSOLORLQIAMPQOQGRPLYTHLFGVIARE"
4938..5597
/gene="B10b"
/feature="B10b"
4938..5597
/gene="B10b"
/codon_start=1
/transl_table=11
/product="gethobiotin synthetase"
/protein_id="AA23518.1"
/db_xref="GI:145428"
/translation="MSKRYEVTGDTVEGKTVASCALLQAAKRAQRYTAGKRPVNSG
EKTEPEGRNSDALALORNSSLDLDYAVVNPYTFAPRPSPHIISAOBGRPTESLMSAG
LRALEQOADMVIVGAGGWFPLSDPTFADMDYQDOLPVLIVYGVKGLGCTINAMLA
QVIQHAGLTLTAGWANDVTPPKRHAEYMTTLIRMLPRCWEKSPGLQKMRQPESS
T"
BASE COUNT      1363 a 1554 c 1631 g 1245 t
ORIGIN      4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match      100.0%; Score 25; DB 2; Length 5793;
Best Local Similarity 100.0%; Pred. No. 5.2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 attgtcgcaagtcacagattattc 25
Db      2035 ATTGTCGCAAGTCACAGATTATT 2059

RESULT 6
LOCUS      A38246      5872 bp      DNA
DEFINITION      Sequence 1 from Patent WO9408023.
ACCESSION      A38246
VERSION      A38246.1 GI:2294844
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM      Escherichia coli.
REFERENCE      1 (bases 1 to 5872)
AUTHORS      Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE      BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL      Patent: WO 9408023-A 1 14-APR-1994;
COMMENT
other publication PL 308301 950724
other publication CA 2145400 940414
other publication AU 4820293 940426
other publication HU 71781 960228
other publication SK 42095 951108
other publication CZ 9500809 950913
other publication FI 951547 950331
other publication JP 8501694T 960227.
location/Qualifiers
1..5872
organism="Escherichia coli"
strain="DSM498"
db_xref="taxon:562"
clone="PB030A-15/9"
1..96
function="PROMOTER PTAC"
evidence=experimental
23..28
-35_signal
-10_signal
45..49
standard_name="PROMOTER PTAC"
evidence=experimental

RBS      105..119
/strand_name="B10b RBS NO.9"
/evidence=experimental
117..1157
/gene="B10b"
117..1157
/gene="B10b"
/codon_start=1
/transl_table=11
/number=1
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAA02325.1"
/db_xref="GI:2294845"
/translation="MAHRRPRLTISOVTELEFEKPLDLLFEAQOVYRHFDRQOVST
LUSIKGACPECKCYCPOSSRYKTGLEPARELMEVQVLESARKAKAAGSTRKMGAAW
KNPHEDMRYLEOMVOGVAKMGLKACMTLGLTSESQAOQLANAGLDYNNHNDLTPES
YGNLITTRVYQRLDTLEKVRDAGIKVSGGIVGIGETVKRAGILLQLANLPPEPES
VSNILNVKVGTPILADNDVDADFIRITAVARIMPTSYRLSAGREOMBOQAMC
FMAGANSIEYGCGLTTPNPEEDKDLQLEFRKLGILNPQOTAVIAGDNQOQRLQALMT
PDTDEYNNAAAL"
2284..2297
/standard_name="B10C RBS"
2295..3050
/gene="B10C"
2295..3050
/gene="B10C"
/function="INVOLVED IN PIMELQYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/number=3
/product="PROTEIN"
/protein_id="CAA02326.1"
/db_xref="GI:2294846"
/translation="MATVAKQAIAAFGRAAAHYEHOADLOROSADALLAMLPORRYT
HYLDAGCGPMMSRHMRERHQAQVATLDSPLMIVQARQKADADHYIAGDIESLPLATA
TFDLAMSNLAVOWCGLNSTALRELRYRVAPKGVYAFPTTVGSLPERHQMQAVDERP
HANRFLPPDEITEOSINQVYOHHTOPTITLWPDDELASMRSLKIGATHLHGRPRIL
TRSOLORLQIAMPQOQGRPLYTHLFGVIARE"
3742..3752
/standard_name="B10A RBS"
3750..5039
/gene="B10A"
3750..5039
/gene="B10A"
/standard_name="S-ADENOSYL-L-METHIONINE:
8-AMINO-7-OXONONANOATE AMINOTRANSE."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/number=5
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAA02327.1"
/db_xref="GI:2294847"
/translation="MTTDLAPDORHIMHPYTSPLPYVPVYSAEGCELITSGSR
LVGMSMMAAIGHYNNPOLNAAKRSIOAMSHVGGITHPADIELCCKVLAMPPO
LECVPLADSGSAVEVAMKALQYMAKGRARQRTTFRNGTHGDTFGMSVCDPDS
MHSIMKCYLPENLFAPOSRRDGEWDERDVGFAFLMAHREHETIAVILIEPVGAG
GMAMYPENIKRIRKICDRBGLILADETATGEGRGLFACHEHETIADITLCKAL
TGGITWLSATLITTRVEAETISNGEACFMHGPFGKPLACAEHETIADITLCKAL
QVADIEVQLREQLAPARDAEMVADVAVLGAIGVETTHPYNNAALOKFVEGVWIRP
FGKLIYIMPPYILLPOOLRLTAAVRAVADDEFQO"
5088..5100
/standard_name="ORF1 RBS"
5098..5574
/gene="ORF1"
5098..5574
/gene="ORF1"
/number=6
/standard_name="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/number=6

```

```

/evidence=experimental
/product="PROTEIN"
/protein_id="CAA02328.1"
/db_xref="GI:2294848"
/translation="MKLSINDLRGDKLPHRRVFMGMGVDGDNISPHLAMPDVPAGTK
SFVAVTCYDPDAITGSGMMHWYVNI.PAOTRVLPQFSGGLVAMPDGVLOTEDCKTKG
YDGAAPKGETHRIIFYVALDIERIDVDEGASGAMVGFNVFHSLASATITAMS"
5583..5605
5583..5644
stem_loop
terminator
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best local Similarity 100.0%; Pred. No. 5.2e+05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 attgtcgcaagtcacagaattatt 25
|||||
Db      140 ATTGTCCAGTCACAGAAATTATT 164

RESULT 7
A38251      5872 bp      DNA      PAT      05-MAR-1997
LOCUS      Sequence 6 from Patent WO9408023.
DEFINITION A38251
ACCESSION  A38251
VERSION     A38251.1 GI:2294849
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM    Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE      BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL     Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694* 960227.
FEATURES
Source      Location/Qualifiers
            1..5872
            /organism="Escherichia coli"
            /strain="DSM498"
            /db_xref="taxon:562"
            /clone="PB030A15-9"
            1141..1156
            /standard_name="BIOF RBS"
            1154..2308
            /gene="BIOF"
            1154..2308
            /gene="BIOF"
            /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
            /EC_number="2.3.1.47"
            /codon_start=1
            /transl_table=11
            /number=2
            /evidence=experimental
            /product="KAPA SYNTHASE"
            /protein_id="CAA02329.1"
            /db_xref="GI:2294850"
            /translation="MSMOEKINALDARRADALRRXYPAVAGCGRWLVADDDROYLNF
            SSNDYLGISHHPQILIRAMQOGAFOFGSGSGVSGIVYHQLLEELAEMLCYSVA
            LLFTSGFAANAVAIAMAKEDRIADRLSHASLLLEASUSPSQLRRPAHNDVTHLAK
            LLASPCGQGMVTEGVFSMDGDSAPLAEITQVYIOHNGWLMVDADAGTGVIGQGGG

```

```

SCWLOKYPPELLVYTFCKGFGVGAAYICSTVADYLLOFARHLIYSTMPAQOAL
RASLAVIRSDGDRREKIALILTRPAGVODLPTLLAOSCSAIOPLIYGDNSRALQL
AEKLRQGCWATLRPTVPAGTARLRLTLAAHEMODDRLLEVLHGNG"
3030..3045
/standard_name="B10D RBS"
3043..3753
/gene="B10D"
3043..3753
/gene="B10D"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="PTB SYNTHASE"
/protein_id="CAA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYEVTGDTVEGKTVASCALLQAAGYRTAGYKPVASGS
EKTEGELNSDALIQRSSSLQDLYATVNPYTFAPESPHISAOGRPTESLVMASG
LRALFQADWYLVGAGGMPPLSDFTFADWVVOEOLPVILYVGKIGCINHMLTA
OYQHAGITLGGWANDVYTPPGKRHAETMTLTMTIPAPLGLIPLWLAENPENATGK
YINLAFVDSYTLGFTSRL"
BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best local Similarity 100.0%; Pred. No. 5.2e+05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 attgtcgcaagtcacagaattatt 25
|||||
Db      140 ATTGTCCAGTCACAGAAATTATT 164

RESULT 8
A93674      5872 bp      DNA      PAT      22-JAN-2000
LOCUS      Sequence 1 from Patent EP0798384.
DEFINITION A93674
ACCESSION  A93674
VERSION     A93674.1 GI:6741862
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM    Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch,O. and Brass,J.
TITLE      Biotechnological method of producing biotin
JOURNAL     Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
FEATURES
Source      Location/Qualifiers
            1..5872
            /organism="Escherichia coli"
            /strain="DSM498"
            /db_xref="taxon:562"
            /clone="PB030A-15/9"
            1..96
            /function="PROMOTER PTAC"
            /evidence=experimental
            23..49
            /evidence=experimental
            45..49
            /standard_name="PROMOTER PTAC"
            /evidence=experimental
            105..119
            /standard_name="B10B RBS NO.9"
            /evidence=experimental
            117..1157
            /gene="B10B"
            117..1157
            /gene="B10B"
            /codon_start=1

```

```

/transl_table=11
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHRPMTLSQVTELEFEKPLDLLFEAQOVRHOFDDPROVOST
LSTIGACPEDECKYCKPOSSRYKTGLEMERLEMEVOYLESRKAKAASTYFCQAMAM
KNPHERMDPELOMVOGVKAMGLEACMTLGLTSEQARLNACIDYTNHMLDTSPEF
VGNITTRTYOERLDTLEKVRAGIKVSGGIVGLGTVKDRAGLLQLANLPPTES
VPIIMLVYKGTPTLADNDVDADFIRITIAVARLMPSTSVYRLSAGREOMEOIQAMC
FMAGANSITFGCKLITTPNPEEDKDLQLEFRKIGLNPOOTAVLAGNEQOORLEQALMT
EDTDEYNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
CDS
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVNVKQALAAFGRAAAHYEQHADLQROSADALLMLPORKYT
HLDAGCGPMWSHMRERHQAQVTALELSPMLVQAKQKADAHYLAGDIESLPIATA
TFDLAMSNLAVQMGCLNSTALRELRYRVRPGVYAFITLVQSGLPETHOMQAVDERP
HANRFLPDEIEOSLNGVHOHHIOPITIMPDALSMRSRLKGTGATHLEGRDPRIL
TRSORLOLAWPQGGRYPLTYHLFLGVLARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE-8-AMINO-7-
OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLAFDRIHWPYTSMTSPLPYVYVSAEGCELLISDGR
LVGMSWMAIIGHNVNHPOLNAAMKSQIDAMSHMFQGITHPALIELCRKIVAMPPO
LECVFLADSGSVAVAMKMLQYMOAKGAPRQRLTPRNGYHGDTEGASNCDDNS
MSLMKGYLPENLEFAPQOSRMDGEMDERMNGFARLMAHREINAVIIEPIVQAG
GMVMYHEMLKRIKRTICDRGILLINDETATGCRGKLPACBHAETADITCLGAL
TGGMTLSATLTTRVAVERTISNEEACCFMHGPTFMGNPLACAAANASIALIESGDMQ
QVADIEVOLREQAPARDAEMVADVLCAGVETTHPVNMAALQKFEEOGVWIRP
EKLITLMPYITLLPOOIORLTAANRAVODETFEQO"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDGDLPKRRHVFNGMGYDGNISPLHAMDVPAGTK
SEVVTCYDPDAPGSGMWMVYVNLPAIDRVLPQGGGIVAMPDGVYIOTTRDQKTS
YDGAAPKGETHRYLFTYHALDIERIDVDEGASCAWCPNHFHSLASASTTAMFS"
5583..5605
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
1552..1695
/gene="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
1552..1695

```

BASE COUNT 1318 a 1552 c 1695 g 1307 t

```

ORIGIN
Query Match
Best Local Similarity 100.0%; Score 25; DB 9; Length 5872;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgtcgcaagtcacagaattattc 25
Db 140 ATTGTCGCAACTCACAGATATTATT 164

RESULT 9
A93679
LOCUS A93679 5872 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 6 from Patent EP0798384.
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 5872)
REFERENCE
1 Birch, O. and Brass, J.
Biotechnological method of producing biotin
Patent: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)
FEATURES
Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSWQEKINAALDARRADALRRRYPVAGAGRWLVADNQYLFN
SSNDYIGSHHPQIIRAMOGGAPQPGSGSGSVSYVHQALEELABWGLYSNA
LLFTSGFPAHQVITAAAMAKEDRIAADRLSHASILEASLSPSOLRRFPAINDYTHLAR
LLASPCQOQWVTVIEGFSMDGSDAPLAETIQOYVQNGMLMDDAGTGVIGEOGRG
SCWLOKAVKDELLVTVTFKRGVSGAAVLCSTVADYLLQFARHLITESTMPAQAOAL
RASLAVIRSDDEGADARRKELIALITRFPAQVODLPFTLADSCSAIOPITVDNSRALQ
AEKLROOCQWVTAIRPPTVPAGTARLRULTLAHEMDDIDRLLEVLHGNC"
3030..3045
/standard_name="BIOB RBS"
3043..3753
/gene="BIOB"
3043..3753
/gene="BIOB"
/gene="BIOB"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRRYVGTDTIEVGKTVASCALLQAAKAGYRTAGYKPVASGS
EKTPBGRNSDALALQNSSLQDVAIVNVPYFAEPTSPHIIISAOGRPIESLWMSAG
LRALQOQADWVIVLEGAGGWTPLSDPTFADWVTOQLPVLIVGVKLCGINHAMLTA
QVIQHAGLTLAGVAVNDVTPPGKRHAIEYVTTLTRMIPAPLLGELPWLAPNPNNAATGK

```


BASE COUNT 1318 a 1552 c 1695 g 1307 t
 YINIAFVASTLIGFRTU"
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 5.2e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atgtcgcaagtcacagaattattt 25
 ||||||||||||||||||
 DB 140 ATTGTCGCAAGTCACAGAAATTATT 164

RESULT 10
 LOCUS AR101809 5872 bp DNA PAT 14-FEB-2001
 DEFINITION Sequence 1 from patent US 6083712.
 ACCESSION AR101809
 VERSION AR101809.1 GI:12812607
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.

REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
 FEATURES Location/Qualifiers
 source 1..5872

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 5.2e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atgtcgcaagtcacagaattattt 25
 ||||||||||||||||||
 DB 140 ATTGTCGCAAGTCACAGAAATTATT 164

RESULT 11
 LOCUS AR101810 5872 bp DNA PAT 14-FEB-2001
 DEFINITION Sequence 6 from patent US 6083712.
 ACCESSION AR101810
 VERSION AR101810.1 GI:12812608
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.

REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
 FEATURES Location/Qualifiers
 source 1..5872

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 5.2e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atgtcgcaagtcacagaattattt 25
 ||||||||||||||||||
 DB 140 ATTGTCGCAAGTCACAGAAATTATT 164

RESULT 12
 LOCUS AE000180

DEFINITION AE000180 11022 bp DNA BCT 01-DEC-2000
 Escherichia coli K12 MG1655 section 70 of 400 of the complete
 genome.

ACCESSION AE000180 U000096
 VERSION AE000180.1 GI:1786988

KEYWORDS

SOURCE

Escherichia coli K12.
 Escherichia coli K12.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE 1 (bases 1 to 11022)
 AUTHORS Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
 Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
 Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
 Mau,B. and Shao,Y.

TITLE The complete genome sequence of Escherichia coli K-12
 JOURNAL Science 277 (5331), 1453-1474 (1997)
 MEDLINE 97426617
 PUBMED 9278503

REFERENCE 2 (bases 1 to 11022)
 AUTHORS Blattner,F.R.

DEFINITION Direct Submission
 Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459

REFERENCE 3 (bases 1 to 11022)
 AUTHORS Blattner,F.R.

DEFINITION Direct Submission
 Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459

REFERENCE 4 (bases 1 to 11022)
 AUTHORS Plunkett,G. III.

DEFINITION Direct Submission
 Submitted (13-OCT-1998) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 This sequence was determined by the E. coli Genome project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCHGR). The entire sequence was independently
 determined from E. coli K12 strain MG1655. Predicted open reading
 frames were determined using Genemark software. Kindly supplied by
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.

30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
 have been correlated with genetic loci are being annotated with CG
 Site Nos., unique ID nos. for the genes in the E. coli Genetic
 Stock Center (CGSC) database at Yale University, kindly supplied by
 Mary Berlyn. A public version of the database is accessible
 (<http://cgsc.biology.yale.edu>). Annotation of the genome is an
 ongoing task whose goal is to make the genome sequence more useful
 by correlating it with other data. Comments to the authors are
 appreciated. Updated information will be available at the E. coli
 Genome Project's World Wide Web site
 (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and
 its annotations are periodically updated; this is version M54. No
 sequence changes. Annotation updates: updated gene identifications
 and products; all new functional assignments courtesy of Monica
 Riley; added promoters, protein binding sites, and repeated
 sequences described in reference 1. The unique numeric identifiers
 beginning with a lowercase 'b' assigned to each gene (protein- or
 RNA-encoding) are now designated as gene synonyms instead of
 labels. This should allow them to be searched for in Entrez as gene
 names.

FEATURES
 source

Location/Qualifiers
 1..11022
 /organism="Escherichia coli K12"
 /strain="K12"
 /sub_strain="MG1655"

```

repeat_region /db_xref="taxon:83333"
               <1. 130
               /note="REP (repetitive extragenic palindromic) element;
               contains 4 REP sequences"
gene           complement(147. 1430)
               /gene="ybhc"
               /note="b0772"
               complement(147. 1430)
               /gene="ybhc"
               /function="orf; Not classified"
               /note="1427; 98 pct identical to fragment YBHC_ECOLI
               SW:P46130 (300 aa) but contains 127 additional C-terminal
               residues"
               /codon_start=1
               /transl_table=11
               /product="putative peclnesterase"
               /protein_id="AAC73859.1"
               /db_xref="GI:1786989"
               /translation="MNTFVSRLALALAFVTLTACSTPPDQRPDQTAFTGSRP
               LSAEQNFDAQHYFASLTPGAAMNPSITLPAPDPVVGPAQTGVTHTTIOAAD
               AAITKRTKROYIAVMPGEYGYVYVPAAPGCTITXGGERIDYVKIGLSIDGMSFA
               DMRDVNRKGYKMPGKPAWYWDSCQSRSSICVLSAYFWSQNGIOLQNTENT
               LGSVDACNHPAVLRTDGOVQINNVAITRONTFFVNSGVONRLETNKPRTLV
               NSYIEGDVIVSGCAVFDNTEFRVYNSRTQEAAYFAPATLSNIYGYFAVNSRN
               AFGEVAVOLGRSLVDANTGVVIRDSAINEGFNTAKPMADAVISNRPAAGNTGSVD
               DNDIQRNLNDNTNVRMWEYNNRGVSKVAEAK"
               complement(1507. 1535)
               /note="factor Sigma70; predicted +1 start at 806574"
               complement(1582. 2058)
               /gene="ybhb"
               /note="b0773"
               complement(1582. 2058)
               /function="orf; Unknown"
               /note="1518; 99 pct identical to YBHB_ECOLI SW: P12994"
               /codon_start=1
               /transl_table=11
               /product="orf; hypothetical protein"
               /protein_id="AAC73860.1"
               /db_xref="GI:1786990"
               /translation="MKLISNDRDGRKLRHRYNGNGYDGNLISPHLAMDVDPACT
               SFVATCYDPAPGSGMHWYVNLPAADRVLPDQSGSGLVAMPDGVLOLRIDPFGTG
               YDGAAPKGETHRYITVYHALDIERIDVDGASGAWGFVHFHSLASASTAMFS"
               complement(2108. 2124)
               /note="central position to predicted promoter:85.5"
               /bound_moiety="Rnas predicted site"
               complement(2117. 3406)
               /gene="bioA"
               /note="b0774"
               complement(2117. 3406)
               /gene="bioA"
               /EC_number="2.6.1.62"
               /function="enzyme; Biosynthesis of cofactors, carriers:
               Biotin"
               /note="1429; 100 pct identical to BIOA_ECOLI SW: P12995"
               /codon_start=1
               /transl_table=11
               /product="7,8-diaminopelargonic acid synthetase"
               /protein_id="AAC73861.1"
               /db_xref="GI:1786991"
               /translation="MTTDLAFDORHIMHPTSMSPILPYVYVSAEGCEILLISGRK
               LVDGSSMAAIIHGYNHDOLNAAKSQIDAMSHVAFGGITTHAPALIELCKRLVAMPPOP
               LECVFLADSGSAVEYAKMALQYMAKEAQRFLTFPNGYHGTFFGAMSCDPDS
               MSLSKGYIPENLFAFAPDSRMDGEMDEDMWGFARLMAAHRHETAAVYIEPIYOGAG
               GKRAHPHWLKRIRKICREGILLIADELATGFGRTGLFACFAHAELADPILICGLKAL
               TGGITMLSATLTTRVAAETISNGEAGCFMHGPTFMGNPLACAAVASALATIESGMOO
               OVADIEVOLREQLAPADAEMVADVIVGATIGVETTHPVNMAALQKRFVEGQWIRP
               FEKLIYIMPPYIILPOOLRLFAAVNRAVNDTEFFGQ"
               complement(2193. 2221)
               /gene="bioA"
               /note="factor Sigma70; predicted +1 start at 807260"
               3411. 3450
               /note="central position to biob promoter: -20"

```

```

protein_bind  /bound_moiety="biob documented site"
               3411. 3450
               /note="central position to predicted promoter: -20;genetic
               evidence for the site"
protein_bind  /bound_moiety="biob predicted site"
               complement(3411. 3450)
               /note="central position to predicted promoter:50"
               /bound_moiety="biob predicted site"
               complement(3411. 3450)
               /note="central position to biob promoter:50"
               /bound_moiety="biob documented site"
               3413. 3441
               /note="factor Sigma70; promoter biob; documented +1 at
               808525"
               complement(3447. 3473)
               /note="factor Sigma70; promoter bioA; documented +1 at
               808515"
gene          /gene="bioB"
               /note="b0775"
               3493. 4533
               /note="b0775"
               3493. 4533
               /gene="bioB"
               /EC_number="2.8.1.-"
               /function="enzyme; Biosynthesis of cofactors, carriers:
               Biotin"
               /note="0346; 99 pct identical to BIOB_ECOLI SW: P12996"
               /codon_start=1
               /transl_table=11
               /product="biotin synthesis, sulfur insertion?"
               /protein_id="AAC73862.1"
               /db_xref="GI:1786992"
               /translation="MAHRPRWTLISQVTLFKEPRLDILFEAQVHRQHFDPQVYST
               LLSITGACPEDEKCYQPSRYKTKLEERMEYQVLESARKAKAASSTPGGAAN
               KNPHERMDIYELQWQVQKAMGLEICMPLGTISEQAOALNAGLDYNNHLDPSER
               YGNITITRTYQERLDLEKVRDAGIKVSGGIVGEGVNRAGLILLOALNLPDES
               VPINMLVYKGTPLADNDVDADFIRITIAVARIMPTSYVRLSAGSRQMGQAMC
               FMAGANSIEYGGCKLITTPNPERDKLOLFKRLGLNPQGTAVLADNEQORLEQALMT
               PPTDEYVNAAL"
               4530. 5684
               /gene="bioF"
               /note="b0776"
               4530. 5684
               /gene="bioF"
               /EC_number="2.3.1.47"
               /function="enzyme; Biosynthesis of cofactors, carriers:
               Biotin"
               /note="0384; 100 pct identical to BIOF_ECOLI SW: P12998"
               /codon_start=1
               /transl_table=11
               /product="8-aminoo-7-oxononanoate synthase"
               /protein_id="AAC73863.1"
               /db_xref="GI:1786993"
               /translation="MSWQEKINAALDARRADALRRRYPYAGAGRWLVADROYLNF
               SSNDYLGISHHPDIIIRAWOGAEOFGIGSGSGHVSQYVHQAEELEAEWGYSA
               LLFTISGAANOAVIAAMAKEDRIIAARLRSIALSLAASLSPQLRFRANDVYTHAR
               LLASPCPGQMWVTGEGVSMDSAPLAELIIOYVTOCHNGMLMDDAAGTGVTEGCRG
               SCWLQKVPFLVITPFGKGCYSAANVLCSTYADYILQFARHLITISTMPPAQDAL

```

```

Query Match      100.0% Score 25; DB 1; Length 11022;
Best Local Similarity 100.0%; Pred. No. 5e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 attgtcgcaagtcacagaattatt 25
   |||||
Db 3516 attgtcgcaagtcacagaattatt 3540

RESULT 13
AE005258 13501 bp DNA BCT 21-MAR-2001
LOCUS Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
DEFINITION of 155.
ACCESSION AE005258 AE005174

```



```

/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55142.1"
/db_xref="GI:12513757"
/translation="MLSPYVNLGCSNLSLRNLTPNNRYLSSVRAAIVSDGAV
KGNRTYVATDNKECYTRESHGCEFNLLHRLGKRGESIKRIEVLNATPVSAV
ERGIYHSNRPLPDYAPPELPSYDNRSLVPCNGKNAVYEDAEADATVLMK
FTTSQSENEATSEVACENQYTGASAEKTYGNNDIIGIRMDKINGESLNLISLPAQ
AEHATYDFRLEQKILFVDTEFNVLDRAKNEFNEIDISYVNSDRSWSSESQIMO
SYHGKDLISVLSKI"
/gene="20990"
/length=7160
/size=7160
/gene="20990"
/feature="orf: Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55143.1"
/db_xref="GI:12513758"
/translation="MRPTSLNVLHQSRRSSMSDIEDLVKASVQWIKNNPOLRF
QGDHNIYQOIEALDKIGSTETGRVLLNAIESISRLKSTVYIHNSSPLGMARD
IDAMNHRGTSDFHCNLAIVEYEGEISVVDPRATYVEHLLHVRNNGELAKES
SRASOKYSPILLLEARTVGLGAFSEVLESENKFHEITGMFRRTSTPDSALIHDDNT
VSIGQOVRHLHPLL"
/complement(7668..8144)
/gene="ybhb"
/note="20992"
/complement(7668..8144)
/gene="ybhb"
/function="orf: Unknown function"
/note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 strain
MG1655: B0773"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLRDGDKLPHRHVNGMGYDGNISPHLMDVPAATK
SFVATCYDPAPRGSGMHWVNLPAIDTRVLPGCGSLVAMPDGLQTRDPGKTG
YDCAAPPKGTHRIPTVHALDVERIDVDEGASGAVGVNHFHSLASASITAMS"
/complement(8203..9492)
/gene="b10A"
/note="20993"
/complement(8203..9492)
CDS
Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 4; 9e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 attgtcgcaagtacagaattattt 25
|||||
Db 9602 attgtcgcaagtacagaattattt 9626

RESULT 14
AP002553 297816 bp DNA BCT 07-MAR-2001
LOCUS DEFINITION AP002553 Bk000007
ACCESSION AP002553.1 GI:13360211
VERSION
KEYWORDS
SOURCE
Escherichia coli O157:H7 (strain:O157:H7, sub-strain:RIMD 0509952)
DNA.
Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (sites)
REFERENCE
1 (sites)

/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55142.1"
/db_xref="GI:12513757"
/translation="MLSPYVNLGCSNLSLRNLTPNNRYLSSVRAAIVSDGAV
KGNRTYVATDNKECYTRESHGCEFNLLHRLGKRGESIKRIEVLNATPVSAV
ERGIYHSNRPLPDYAPPELPSYDNRSLVPCNGKNAVYEDAEADATVLMK
FTTSQSENEATSEVACENQYTGASAEKTYGNNDIIGIRMDKINGESLNLISLPAQ
AEHATYDFRLEQKILFVDTEFNVLDRAKNEFNEIDISYVNSDRSWSSESQIMO
SYHGKDLISVLSKI"
/gene="20990"
/length=7160
/size=7160
/gene="20990"
/feature="orf: Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55143.1"
/db_xref="GI:12513758"
/translation="MRPTSLNVLHQSRRSSMSDIEDLVKASVQWIKNNPOLRF
QGDHNIYQOIEALDKIGSTETGRVLLNAIESISRLKSTVYIHNSSPLGMARD
IDAMNHRGTSDFHCNLAIVEYEGEISVVDPRATYVEHLLHVRNNGELAKES
SRASOKYSPILLLEARTVGLGAFSEVLESENKFHEITGMFRRTSTPDSALIHDDNT
VSIGQOVRHLHPLL"
/complement(7668..8144)
/gene="ybhb"
/note="20992"
/complement(7668..8144)
/gene="ybhb"
/function="orf: Unknown function"
/note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 strain
MG1655: B0773"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLRDGDKLPHRHVNGMGYDGNISPHLMDVPAATK
SFVATCYDPAPRGSGMHWVNLPAIDTRVLPGCGSLVAMPDGLQTRDPGKTG
YDCAAPPKGTHRIPTVHALDVERIDVDEGASGAVGVNHFHSLASASITAMS"
/complement(8203..9492)
/gene="b10A"
/note="20993"
/complement(8203..9492)
CDS
Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 4; 9e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 attgtcgcaagtacagaattattt 25
|||||
Db 9602 attgtcgcaagtacagaattattt 9626

AUTHORS
Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C. and Shinagawa,H.
Complete nucleotide sequence of the prophage VP2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
2 (sites)
Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
Hayashi,T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
3 (sites)
Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
Shinagawa,H.
Complete nucleotide sequence of the prophage VP1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
4 (sites)
Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C.-G., Ohnishi,E., Nakayama,K., Murata,T.,
Tanaka,M., Tode,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
Shinagawa,H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
21156231
5 (bases 1 to 297816)
Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
Hayashi,T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center: 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail: ken@gen-1nfo.osaka-u.ac.jp),
URL:http://www.gen-1nfo.osaka-u.ac.jp/, Tel:81-6-6879-8365,
Fax:81-6-6879-2047)
genome project.
location/Qualifiers
1..297816
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
79..1245
/gene="ECs0753"
79..1245
/gene="ECs0753"
/note="ECs0753"
/note="similar to SUCC_ECOLI g11786948 percent identity
100 in 388 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/feature="not experimental"
/product="succinyl-CoA synthetase beta subunit"
/protein_id="BAB34176.1"
/db_xref="GI:13360212"
/translation="NNLHEYOAKQLFARYGLPAPVGYACTTPPEAEASKIGAGPW
VKCOVHAGGKAGKAGVKNVSKEDIPAFENMLGKRLVYQDANQOPVNOITVEAT
DIAKELYLEGAVVDRSSRRVFNASFTGVEIKVAEETHLHKVALDPLTPMPYQG
RELAFKLYEGKLYVOOFTKIFMKLATIFERDLALIEPLVITKQGLICIDKQIGA
DGNALFPROPLIREMRQSDPREAOAGWELNVALDNLGICVNGAGLAIAGTMDIY
KLHGEPANFLVDGGATKERVYAEKRIILSDKRVAVNLVETGIVRCDLTDADIIG
AAVEGVNVPVVRLEGNAELGAKKLADSGINIIAAKGLTDAOOVVAABVGRK"

```

gene
1245. .2114
/gene="ECS0754"
CDS
/note="similar to SUCC_ECOLI g11786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/translation="MSILIDKNTKVICGFTSGCTFHSEQAINGTKMVGVTGKG
GTHLGLPENTVREAVATGATASVIVYAPPCDSILLEAIDAGIKLITTEGPT
LDMLIVKVLDEAGVBMIGPCGVITPGECKIGIOPHHPKGVISRSGLTLYE
AVKOTDYGSGTGVIGIDPISGNSFIDILEMPEKDPQTEAIVMIGIGSAESEA
AAYIKHNTKPVVGYTAGVTAPEKGMHAGAI IAGKGTADKEKFALEAGVKTIRS
LADIGFALKTVLK"
2781. .3686
/gene="ECS0755"
CDS
/note="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
transcription activator (gcv operon activator)
Escherichia coli g14170431sp1P320641GCVL_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"
/translation="MRGKIPKTELIVFEVVAHESYTRAEEELATQSAFPAQVSL
EELRPPLFSHKKRIPLNDTGKYLIVGIVETINKLERDNTIMTQPTQVLELVN
PPESTHMLIPNHEPTKLPDIIIVNHSILANNDDREYDAVIMENFCAPAVEY
LPREETLIPVSGSLAMDOKLSVAELITPELPHOSTRITGMEEMFALSGVSPLVN
NGPRELDLSMLIAVNSNLGVALLPRAITDHDSDGMVTPCVPPIRTGNRTIMTQE
EKSDSPHLQOFREWLAKSVVQDM"
complement(3720. .4322)
/gene="ECS0756"
CDS
/note="probable cob(I)alamin adenosyltransferase, similar
to cob(I)alamin adenosyltransferases (corrinoid
adenosyltransferases) e.g. [Escherichia coli]
g1151481sp1P130401BTUR_ECOLI percent identity 67 in 200
aa"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="putative cob(I)alamin adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/translation="MEARISTEPIHVOHQOKLEQVDFRVAALIEKKGILIVETNGKG
KSTAAFGTVRAVGHGKTVGAQYIKQMDGNEINLIQPCVEFHNGTGTMTGNR
QADIDAAKEVSESKRMADKRYDLVDLDELTLATYHIDTTEVIASLQNRPAQOSY
IVTGRGCHSOLIKMADYSEIRPVKHAFFDNGIQAPGIDW"
complement(4332. .5984)
/gene="ECS0757"
CDS
/note="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
(fumarase) - Escherichia coli
g11205981sp1P009231PUMA_ECOLI percent identity 68 in 545
aa"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"

gene
/translation="MSKPIIOWEPFLONKDGTEXTLISDHIITVELDGEVYKIAPE
ALTLIARQATFEASFLPRLSARHIOVASTILINDQOASSNDKVVALLRLMAEVSAYLP
NCODTGTATVYASGQOIWIGNDNAEHLASKYITSPENNLRSQNALPDMYREVMTO
TNLPAQIDISAVNADDEYHFLCVNKGSSANALVYOTKSLDPEKITALIEKMSL
GTACAPRYHIAFVVGGLSADOTLRIALASTKYIDNLPISGNEGQAFROIIEKYL
EASOFGIGAGGKGFADHRIYRLPRHSGSCP IAMALSCSADRNKAIKIRHGMVL
KLEHNPQGYIPASLREHNOAHQVOLDNRLRVMODLARLPVGTAVSLSGPVYAR
DIARAKTARLDGSEPMPEYKHHIIVYAGPARKPEMAGCSIGPTGGRMDGYIDFE
QAAAGSLVWLSKGRNSQOYTDACKRHGSEFNIGSGAALLAOLYVKSLECLEPELGG
MEAVWMMMEVENLPALVDDKGNFESOFEDQHCASCAPAGH"
complement(6092. .7372)
/gene="ECS0758"
CDS
/note="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. (Bacillus stearothermophilus)
g11214671sp1P249431GUTR_BACST percent identity 38 in 416
aa, also similar to C4-dicarboxylate transport"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translation="MKKISLTMTMLALVLMIGVNLNNTASPEAKLYAOEISIFP
TFPLIKIMLIAPVYVSTLVYVIAKMGAKALGRFSTLFLFICASLSIALGLIVY
NFFMPGTGINFAHGAETTYAAPEPTLPKFIASHAPTSIYDAMAHNEIIOVESI
PLGCSILTLAGEGSAIVHALDSLAAHMLVGVNLPVYFAAISLALBRLGAVP
VAGSIPMGEEFFYPTMLLVLLIGLIVVGPCTRLRLTRALEPALATTSSEARVP
GTLEKLEQSGVSPKIASFVLPYIGSFNLSMAYCSFATVFAQCNTHLSIGQITM
LILMLTSGMAGVPRASVNVYIAATLNFNIEAQLILMLGVDPFLDMGRSATVMSN
AAGAMVSRWEGEHGEGCGKALKPESNVALLP"
complement(7533. .7850)
/gene="ECS0759"
CDS
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"

Query Match 100.0%; Score 25; DB 2; Length 297816;
Best Local Similarity 100.0%; Pred No. 3; 9e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgtcgcaagtcacagaattatt 25
Db 92334 ATTGTCCGACAGTCACAGATTATT 92358
|||||
RESULT 15
AF250776 DNA BCT 31-JAN-2001
LOCUS 5526 bp
DEFINITION Uncultured bacterium pCOSHE2 hypotheical 17.1 kDa protein in
modc-biocl intergenic region, DAPA-amino transferase Biocl (biocl),
biocl synthase Biocl (biocl), KARA synthetase Biocl (biocl), and
biotin biosynthesis protein Biocl (biocl) genes, complete cds; and
dehliobiotin synthetase Biocl (biocl) gene, partial cds.
ACCESSION AF250776
VERSION AF250776.1 GI:12620124
KEYWORDS uncultured bacterium pCOSHE2.
SOURCE uncultured bacterium pCOSHE2.
ORGANISM Bacteria; environmental samples.
REFERENCE 1 (bases 1 to 5526)
AUTHORS Entcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streif, W.R.
TITLE Direct cloning from enrichment cultures, a reliable strategy for
isolation of complete operons and genes from microbial consortia
JOURNAL Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
MEDLINE 20575196

PUBMED 11133432
 REFERENCE 2 (bases 1 to 5526)
 AUTHORS Entcheva, P., Liebl, W. and Streif, W.R.
 TITLE Direct Submission
 JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
 Goettingen, Griesbachstr. 8, Goettingen 37077, Germany
 FEATURES
 source
 1.5526
 /organism="uncultured bacterium pCosHE2"
 /db_xref="taxon:143797"
 /clone="pCosHE2"
 /note="unknown organism, cosmid clone derived from
 environmental consortium"
 complement(52.528)
 /note="ORF1"
 /codon_start=1
 /transl_table=11
 /product="hypothetical 17.1 kDa protein in modc-b10A
 intergenic region"
 /protein_id="AAG60577.1"
 /db_xref="GI:12620125"
 /translation="MKLISNDLRDGLKPHRHVNGMGYDGDNISPHLANDVYDPAKTR
 SFVVCYDPAKTRSSGMWVYVNLPAIDTRVLPQFGSGGLVAMPDGLYRTDEGKAG
 YDGAAPPKGETHRTFTVHALDIERIDVDEGASGAVGVNHFHSLASATFMFS"
 complement(587.1876)
 /gene="b10A"
 complement(587.1876)
 /note="7,8-diaminopelargonic acid
 synthetase-aminotransferase"
 /codon_start=1
 /transl_table=11
 /product="DAPA-aminotransferase b10A"
 /protein_id="AAG60578.1"
 /db_xref="GI:12620126"
 /translation="MTTDLAFDQRIHWPYTSMTSPLPVYVSAEGCELLISDGR
 LVDSMSWMAIHGYNHPQLAAKRSQIDMSHVFGIITHAIELCRKLVAMTPQ
 KLCVPLADSGSVAEAMKMAIOWAKGARGRFLFRNGYHGDFFGASVCDPNS
 MSLMKGYLPENLFAFAPQSRMDGEMDERPVEFARLMAHREIAVIAIIEPVQAG
 GKNVHPEWMLKRIKICDREGILADIETAGRGRIKIFACSHATIAIDICLAL
 TGGMTLSATLITREVAETISNGEACGFMGPMPKPIACAAANSLALISGDWQ
 QVADIEVQLEQALAPARADAVDYVLCIGVETTHPVNMAALQKFEVEQGVWIR
 FGLTIYMPRTIILPQOLRLTAANNAVQDEFECQ"
 1963.3003
 /gene="b10B"
 1963.3003
 /gene="b10B"
 /codon_start=1
 /transl_table=11
 /product="biotin synthase b10B"
 /protein_id="AAG60579.1"
 /db_xref="GI:12620127"
 /translation="MAHRRWTLISQVTELEFEKPLDILFEAOVHROHDFPROVOYT
 LLSIKTGACPEDECKYCPOSSRYKTEGARIMEVGVLESARAKAASGTRCMAAM
 KNNERDMPLYLEOMGVKDIIEACMTGTSISQAOLANAGLITNHNLDISPEF
 YGILITRTYQERDITLEKVRAGITVCGGIVGGEIVKDAAGLLDLQANLPTPEP
 VRIENMLVYKGTPLANDVDVADFDITRTIAVARIMPVSYRISGREDQMNQOTQAMC
 FNAAGSIFVGGCLITTPNEEDKDLQIRKLGILNPQOTAVLAGDNEQOORLEQALMT
 PTDDEYNAAL"
 3000.4154
 /gene="b10F"
 3000.4154
 /gene="b10F"
 /note="8-amino-7-oxononanoate synthase"
 /codon_start=1
 /transl_table=11
 /product="KAPA synthetase b10F"
 /protein_id="AAG60580.1"
 /db_xref="GI:12620128"
 /translation="MSWCKIINALIDARRADALRRYPVAGAGRWLVADROYLNF
 SSNDYLGSHHPQITIRAWQGAEGFSGSGSVAGSVVAHLEELAEALWLGYSRA
 LFTISGFANQAVIAAMAKEDRIADARLSHSLAEASLSQSLRRRAHNDITLAR
 ILASPCGQOLVVTGCVSMGDSAPLAEIQVTVQHNHGMWLVDAHGTGYIGEGRC

SCMLQKPELVVTFGKSGVSAVICSSTVADYLILOFARHLIYSTMPAQAOL
 RASLAVIRREDSBARKKSLAHFRAGVODPFLADSCSHIOPILVGDNSRALOL
 AKLRQCCVYTAIRPPVYAGTARLRLTLTAHEMODIRLLEVLHNG"
 4141.4896
 /gene="b10C"
 4141.4896
 /gene="b10C"
 /note="reaction step prior to synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=11
 /product="biotin biosynthesis protein b10C"
 /protein_id="AAG60581.1"
 /db_xref="GI:12620129"
 /translation="MATVNRKQALIAAFGRALYRQHADIOROSADVLLAMPKRT
 HYLDACGPGMMTRHWRERHAYTALDLSPPHLYAROKDADHDLADISPLATA
 TEDLANSNAVOMCGULSTRALRHLRVNSKSVAFPTLVGSDLELQAWQVDERP
 HANRFLPEIDSLGVHYOHRIQPTILMFDDALSMRSLKIGATHLHRCRPRIL
 TNSQLRLQLAMPQOGRVPLTYHFLGVARE"
 4883.5526
 /gene="b10D"
 4883.5526
 /gene="b10D"
 /note="DTP synthetase"
 /codon_start=1
 /transl_table=11
 /product="dethiobiotin synthetase b10D"
 /protein_id="AAG60582.1"
 /db_xref="GI:12620130"
 /translation="MLVSKRYVPTDTEVNGKYASCALLQAAKAGRTGYRPAVS
 GSEKPEGRNSDALAIORNSIOLDYATVNPYTFAPTSPIHISAGKRIESSVMS
 SCLRALBOADVYLVGAGGKFTPLSDFTFADWVTOBOLVPLVIVGKGLCINHAML
 TQAOIHGGLTVLAGVANDVPEPKRAEYITTLRMIPALDELIPWLAENPE"
 1274 a 1507 c 1567 g 1178 t

BASE COUNT
 ORIGIN

Query Match
 Best Local Similarity 100.0%, Pred. No. 0.00021;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 tttgcgaagtcacagaattatc 25
 |||||||||||||||||||||
 Db 1987 ttgtccgaagtcacagaattatt 2010

RESULT 16
 AP000954/c DNA 27-JAN-2000
 LOCUS Homo sapiens genomic DNA, chromosome 21q21.1-q21.2 clone:B705D10,
 L156-APP region, complete sequence.
 ACCESSION AP000954
 VERSION AP000954.2 GI:6778729
 KEYWORDS Htg.
 SOURCE Homo sapiens DNA, clone:B705D10.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 25256)
 Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
 Fujiyama, A., Yada, T., Totok, Y., Watanabe, H. and Sakaki, Y.
 Homo sapiens 25,256 genomic DNA of 21q21.1-q21.2
 Published Only in Database (1999) In press
 2 (bases 1 to 25256)
 Fujiyama, A., Yada, T., Totok, Y., Watanabe, H. and Sakaki, Y.
 Direct Submission
 Submitted (17-DEC-1999) to the DDBJ/EMBL/Genbank databases.
 Masahira Hattori, The Institute of Physical and Chemical Research
 (RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1
 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
 (E-mail:hattori@qsc.riken.go.jp URL:http://hgp.qsc.riken.go.jp/
 Tel:81-42-778-9923, fax:81-42-778-9924)
 On Jan 27, 2000 this sequence version replaced g1:6630654.

COMMENT
 JOURNAL

```

Sequence updated (26-Jan-2000).
FEATURES
  Location/Qualifiers
    1..25256
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="21"
    /clone="B705D10"
    /map="21q21.1-q21.2"
BASE COUNT      8452 a      4090 c      4292 g      8422 t
ORIGIN
Query Match      68.0%; Score 17; DB 91; Length 25256;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      9 aagtcacgaattattt 25
        |||
        1714 AAGTCACGAATTATT 1698

RESULT 17
AC087305/c      32767 bp      DNA      HTG      23-DEC-2000
LOCUS      AC087305/c
DEFINITION      Homo sapiens chromosome 18 clone RP11-238P13 map 18, LOW-PASS
SEQUENCE SAMPLING.
AC087305
VERSION      AC087305.1 GI:11990696
KEYWORDS      HTG; HTGS_PHASE0.
SOURCE      human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  AUTHORS      Birren, B., Linton, L., Nusbaum, C. and Lander, E.
  TITLE      1 (bases 1 to 32767)
  JOURNAL      Homo sapiens chromosome 18, clone RP11-238P13
  COMMENT      Unpublished
              2 (bases 1 to 32767)
              Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
              Barta, N., Bastien, V., Boguslavsky, L., Boukhaltier, B., Brown, A.,
              Camarata, J., Campolino, A., Choepel, Y., Colangelo, M., Collins, S.,
              Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S.,
              Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,
              Gargyba, S., Ginde, S., Goyette, M., Graham, W., Hille, I., Johnson, N.,
              Hedges, B., Heaford, A., Horton, L., Hulme, W., Hult, I., Johnson, R.,
              Jones, C., Karatas, A., LaRoque, K., Lamazares, R., Landers, T.,
              Lehoczy, J., Levine, R., Liu, G., Maclean, C., MacDonald, P.,
              Margolis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
              McPheeters, R., Meidrim, J., Meneus, L., Minova, T., Mlenga, V.,
              Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H.,
              O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
              Plunkharg, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
              Ribbeck, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roselli, M.,
              Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P.,
              Sougnez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
              Straus, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
              Travels, M., Travis, N., Triggillo, J., Vassiliev, H., Videl, R., Vo, A.,
              Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,
              Zembek, L., Zimmer, A. and Zody, M.
  DIRECT SUBMISSION
  Submitted (23-DEC-2000) Whitehead Institute/MIT Center for Genome
  Research, 320 Charles Street, Cambridge, MA 02141, USA
  All repeats were identified using RepeatMasker:
  Smit, A.F.A. & Green, P. (1996-1997)
  http://ftp.genome.washington.edu/RM/RepeatMasker.html
  ----- Genome Center
  Center: Whitehead Institute/ MIT Center for Genome Research
  Center code: WITB
  Web site: http://www-seq.wi.mit.edu
  Contact: sequence_submissions@genome.wi.mit.edu
  ----- Project Information
  Center project name: L12071
  Center clone name: 238_P_13

```

```

***** NOTE: This record contains 40 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 707: contig of 707 bp in length
708 807: gap of 100 bp
808 1527: contig of 720 bp in length
1528 1627: gap of 100 bp
1628 2348: contig of 721 bp in length
2349 2448: gap of 100 bp
2449 3157: contig of 709 bp in length
3158 3257: gap of 100 bp
3258 3973: contig of 716 bp in length
3974 4073: gap of 100 bp
4074 4799: contig of 726 bp in length
4800 4899: gap of 100 bp
4900 5617: contig of 718 bp in length
5618 5717: gap of 100 bp
5718 6450: contig of 733 bp in length
6451 6550: gap of 100 bp
6551 7253: contig of 709 bp in length
7260 7359: gap of 100 bp
7360 8066: contig of 707 bp in length
8067 8166: gap of 100 bp
8167 8907: contig of 741 bp in length
8908 9007: gap of 100 bp
9008 9716: contig of 709 bp in length
9717 9816: gap of 100 bp
9817 10532: contig of 716 bp in length
10533 10632: gap of 100 bp
10633 11354: contig of 722 bp in length
11355 11454: gap of 100 bp
11455 12166: contig of 712 bp in length
12167 12266: gap of 100 bp
12267 12976: contig of 710 bp in length
12977 13076: gap of 100 bp
13077 13789: contig of 713 bp in length
13790 13889: gap of 100 bp
13890 14617: contig of 728 bp in length
14618 14717: gap of 100 bp
14718 15452: contig of 735 bp in length
15453 15552: gap of 100 bp
15553 16276: contig of 724 bp in length
16277 16376: gap of 100 bp
16377 17111: contig of 735 bp in length
17112 17211: gap of 100 bp
17212 17925: contig of 714 bp in length
17926 18025: gap of 100 bp
18026 18746: contig of 721 bp in length
18747 18846: gap of 100 bp
18847 19569: contig of 723 bp in length
19570 19669: gap of 100 bp
19670 20384: contig of 715 bp in length
20385 20484: gap of 100 bp
20485 21190: contig of 706 bp in length
21191 21290: gap of 100 bp
21291 22028: contig of 738 bp in length
22029 22128: gap of 100 bp
22129 22841: contig of 713 bp in length
22842 22941: gap of 100 bp
22942 23653: contig of 712 bp in length
23654 23753: gap of 100 bp
23754 24478: contig of 725 bp in length
24479 24578: gap of 100 bp
24579 25305: contig of 727 bp in length

```



```

* 25306 25405: gap of 100 bp
* 25406 26143: contig of 736 bp in length
* 26144 26243: gap of 100 bp
* 26244 26968: contig of 725 bp in length
* 26969 27068: gap of 100 bp
* 27069 27785: contig of 717 bp in length
* 27786 27885: gap of 100 bp
* 27886 28584: contig of 699 bp in length
* 28585 28684: gap of 100 bp
* 28685 29467: contig of 783 bp in length
* 29468 29567: gap of 100 bp
* 29568 30292: contig of 725 bp in length
* 30293 30392: gap of 100 bp
* 30393 31117: contig of 725 bp in length
* 31118 31217: gap of 100 bp
* 31218 31944: contig of 727 bp in length
* 31945 32044: gap of 100 bp
* 32045 32767: contig of 723 bp in length.
Location/Qualifiers
1. 32767
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-238P13"
/clone_id="RPC1-11 Human Male BAC"
BASE COUNT 8045 a 5670 c 5396 g 9330 t 4326 others
ORIGIN
Query Match 68.0%; Score 17; DB 77; Length 32767;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 gcaagtcacacaattat 23
|||||
Db 30730 GCAAGTCACAGATTAT 30714

RESULT 18
AC011937 60072 bp DNA HTG 13-JUL-2000
LOCUS Homo sapiens clone RP11-16E13, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC011937
AC011937.2 GI:7144927
VERSION HTG; HTGS_PHASE0.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 60072)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 60072)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens clone RP11-16E13
JOURNAL Unpublished
REFERENCE 1 (bases 1 to 60072)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collins,M., DeLuca,A.,
Cooke,P., DeLuca,A., DeLuca,M., Dewar,K., Dominko,T., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Haggis,B., Heath,A., Horton,L.,
Howard,J.C., Johnson,R., Jones,C., Kamm,L., Karas,A., Klein,J.,
Lehoczky,J., Lien,C., Locke,K., MacDonald,P., Marquis,N.,
McEwan,P., McQuar,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stegeman,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessier,S., Tjelle,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 3, 2000 this sequence version replaced gi:6056246.

```

```

All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RV/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3522
Center clone name: 16_E_13
-----
* NOTE: This record contains 68 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
1
764 863: gap of 100 bp
864 1659: contig of 796 bp in length
1660 1759: gap of 100 bp
1760 2606: contig of 847 bp in length
2607 2706: gap of 100 bp
3495 3594: gap of 100 bp
3595 4391: contig of 797 bp in length
4392 4491: gap of 100 bp
4492 5168: contig of 677 bp in length
5169 5268: gap of 100 bp
5269 6058: contig of 790 bp in length
6059 6158: gap of 100 bp
6159 6968: contig of 810 bp in length
6969 7068: gap of 100 bp
7069 7865: contig of 797 bp in length
7866 7965: gap of 100 bp
7966 8753: contig of 788 bp in length
8754 8853: gap of 100 bp
8854 9663: contig of 810 bp in length
9664 9763: gap of 100 bp
9764 10563: contig of 800 bp in length
10564 10663: gap of 100 bp
10664 11465: contig of 802 bp in length
11466 11565: gap of 100 bp
11566 12365: contig of 800 bp in length
12366 12465: gap of 100 bp
12466 13274: contig of 809 bp in length
13275 13374: gap of 100 bp
13375 14163: contig of 789 bp in length
14164 14263: gap of 100 bp
14264 15064: contig of 801 bp in length
15065 15164: gap of 100 bp
15165 15965: contig of 801 bp in length
15966 16065: gap of 100 bp
16066 16852: contig of 787 bp in length
16853 16952: gap of 100 bp
16953 17775: contig of 823 bp in length
17776 17875: gap of 100 bp
17876 18674: contig of 799 bp in length
18675 18774: gap of 100 bp
18775 19477: contig of 703 bp in length
19478 19577: gap of 100 bp
19578 20375: contig of 798 bp in length
20376 20475: gap of 100 bp
20476 21252: contig of 777 bp in length
21253 21352: gap of 100 bp
21353 22153: contig of 801 bp in length
22154 22253: gap of 100 bp

```



```

* 54779 54878: gap of 100 bp
* 54879 55683: contig of 805 bp in length
* 55684 55783: gap of 100 bp
* 55784 56484: contig of 701 bp in length
* 56485 56584: gap of 100 bp
* 56585 57369: contig of 785 bp in length
* 57370 57469: gap of 100 bp
* 57470 58276: contig of 807 bp in length
* 58277 58376: gap of 100 bp
* 58377 59169: contig of 793 bp in length
* 59170 59269: gap of 100 bp
* 59270 60072: contig of 803 bp in length.
*
location/Qualifiers
1..60072
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-16E13"
/clone_lib="RPC1-11 Human Male BAC"
15923 a 9785 c 10180 g 17174 t 7010 others
BASE COUNT
ORIGIN

Query Match
Best Local Similarity 68.0%; Score 17; DB 62; Length 60072;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 aagtcacagaattattt 25
|||||
Db 9213 AAGTCACAGAAATTATTT 9229

RESULT 19
HS54J18/C
LOCUS
DEFINITION Homo sapiens chromosome 21 clone RPC1P704J1854 map 21q21. ***
SEQUENCING IN PROGRESS ***, 18 unordered pieces.
AL133366
AL133366.2 GI:10432568
HUG: HTGS_PHASE1.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 86757)
Nordtschke,G., Conrad,A., Dose,S., Grimm,M., Hildmann,T.,
Hornischer,K., Loehnerl,T.H., Scharfe,M., Schoen,O., Yaspo,M.L. and
Bloecker,H.
Direct Submission
Submitted (03-DEC-1999) GBF, Dept. of Genome Analysis, Mascheroder
Weg 1, D-38124 Braunschweig, Germany, E-mail: hornischer@gbf.de,
On Sep 29, 2000 this sequence version replaced gi:6981908.
Collaborators:
GBF, Dept. of Genome Analysis
Mascheroder Weg 1, D-38124 Braunschweig, Germany
and
Max-Planck-Institut fuer Molekulare Genetik
Imnestrasse 73, D-14195 Berlin-Dahlem, Germany
All annotations in this database entry are developed by
computational tools. It is therefore not explicitly noted in the
feature lines that evidence is not experimental.
PROGRAMS AND PARAMETERS USED FOR ANNOTATION:
+++++
+ Analysis and annotation were performed with the automatic
+ 'first-pass' annotation and submission tool
+ 'Annomlter' (Hornischer & Bloecker).
+ Programs used by 'Annomlter':
+++++
+ 'Tandem Repeats': GDE 2.2 option 'tandem'
+ Minimum length 2 bp; Maximum length 20 bp; Score threshold 20
+ Treat N's as mismatches? YES; Allow uniform consensus? NO >
+ 'Inverted Repeats': GDE 2.2 option 'inverted'
+ 'Micro satellites': GDE 2.2 option 'spunkit' (AbaJian) > 'Cpg

```


repeat_region /rpl_family="aaaaaaaaa repeat"
/rpl_type=TANDEM
48093..48128
/note="homology = 83.3%, counts = 18"
/rpl_family="tc repeat"
/rpl_type=TANDEM
48328..48366
/note="IR6, 82% complementary to IR6' (49401..49440)"
/rpl_type=INVERTED
49049..49084
/note="homology = 88.9%, counts = 9"
/rpl_family="aat repeat"
/rpl_type=TANDEM
49049..49069
satellite

Query Match 68.0%; Score 17; DB 84; Length 86757;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 aagtcacagattattt 25
|||||
Db 65420 AAGTCACAGATTATT 65404

RESULT 20
AL353582/c 115551 bp DNA PRI 22-MAR-2001
LOCUS
DEFINITION Human DNA sequence from clone CTD-2190D5 on chromosome 20 contains
ESTs, STSs and GSSs, complete sequence.
ACCESSION AL353582
VERSION AL353582.16 GI:10443412
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 115551)
AUTHORS Skuce, C.
TITLE Direct Submission
JOURNAL Submitted (14-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Oct 1, 2000 this sequence version replaced gi:9988285.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPeP; Information
on the WormPeP database can be found at
http://www.sanger.ac.uk/Projects/Celegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
CTD-2190D5 is from the Caltech genomic sperm BAC library D. VECTOR:
pBeloBAC11

IMPORTANT: This sequence is not the entire insert of clone
CTD-2190D5. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP4-568C11 is at 115452 in this
sequence. The true right end of clone RP5-966J20 is at 100 in this
sequence. This sequence was finished as follows unless otherwise
noted: all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.
Location/Qualifiers

source
1..115551
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="CTD-2190D5"
/clone_11b="CIT-HSP-D1"
1..260
/note="10 copies 26 mer 88% conserved"
1400..1691
/note="Alus9 repeat: matches 1..292 of consensus"
1692..1818
/note="L1PA16 repeat: matches 6020..6157 of consensus"
1878..2076
/note="L1MC4 repeat: matches 7699..7977 of consensus"
2719..3328
/note="305 copies 2 mer cc 52% conserved"
2720..2777
/note="2 copies 29 mer 94% conserved"
2905..3078
/note="2 copies 87 mer 92% conserved"
2955..3494
/note="6 copies 90 mer 85% conserved"
2985..3484
/note="50 copies 10 mer ccccccctctg 60% conserved"
3001..3510
/note="17 copies 30 mer 89% conserved"
3496..3625
/note="13 copies 10 mer cccctctctg 63% conserved"
3512..3631
/note="4 copies 30 mer 87% conserved"
4283..4377
/note="MER5A repeat: matches 63..167 of consensus"
4459..4588
/note="65 copies 2 mer tg 77% conserved"
4460..4579
/note="4 copies 30 mer 80% conserved"
4591..4800
/note="7 copies 30 mer 77% conserved"
4592..4861
/note="3 copies 90 mer 71% conserved"
4619..4800
/note="7 copies 26 mer 79% conserved"
8042..8542
/note="LTR45 repeat: matches 1..513 of consensus"
8945..9067
/note="MER5B repeat: matches 53..177 of consensus"
10365..10706
/note="12 repeat: matches 2350..2674 of consensus"
10990..11142
/note="LTRIG repeat: matches 121..283 of consensus"
11228..11322
/note="LTRIE repeat: matches 468..566 of consensus"
12021..12363
/note="MER46B repeat: matches 1..236 of consensus"
13655..13678
/note="12 copies 2 mer ag 95% conserved"
13942..14438
/note="L1MB5 repeat: matches 5679..6174 of consensus"
15351..15666
/note="2 copies 158 mer 88% conserved"
complement(15754..16330)
/note="match: GSS: Em: A0394110"
15772..16057
/note="LTRIE repeat: matches 267..567 of consensus"
16283..16553
/note="LTRIA1 repeat: matches 1..352 of consensus"
16889..17056
/note="MIR repeat: matches 59..229 of consensus"
17607..17725
/note="Alus9/x repeat: matches 186..302 of consensus"
18349..18558
/note="21 copies 10 mer atacacacat 68% conserved"
18353..18552
repeat_region

```

repeat_region /note="100 copies 2 mer at 62% conserved"
18357..18564
repeat_region /note="8 copies 26 mer 69% conserved"
18358..18567
misc_feature /note="7 copies 30 mer 68% conserved"
/note="complement(18438..18827)"
/note="match: GSS: Em:AQ581588"
complement(18532..18952)
misc_feature /note="match: GSS: Em:AQ633379"
18813..18944
repeat_region /note="PLAM_A repeat: matches 1..133 of consensus"
complement(19334..19731)
misc_feature /note="match: GSS: Em:AQ125259"
19472..19617
repeat_region /note="MIR repeat: matches 78..218 of consensus"
20285..20513
repeat_region /note="MTTD repeat: matches 320..501 of consensus"
20695..20832
repeat_region /note="MIR repeat: matches 53..189 of consensus"
20924..21387
repeat_region /note="MTTD repeat: matches 2..505 of consensus"
21818..21867
repeat_region /note="MER91A repeat: matches 1..48 of consensus"
22143..22304
repeat_region /note="MIR repeat: matches 31..192 of consensus"
22339..22584
repeat_region /note="MTT1A2 repeat: matches 1..225 of consensus"
22585..23315
repeat_region /note="LTR39 repeat: matches 1..794 of consensus"
23316..23455
repeat_region /note="MTT1A2 repeat: matches 225..370 of consensus"
24413..24566
repeat_region /note="MIR repeat: matches 29..200 of consensus"
24745..25217
misc_feature /note="match: GSS: Em:AQ403530"
24758..25162
misc_feature /note="match: GSS: Em:AQ435639"
25811..25956
repeat_region /note="MIR repeat: matches 54..232 of consensus"
27482..27631
repeat_region /note="5 copies 30 mer 77% conserved"
27497..27631
repeat_region /note="9 copies 15 mer 79% conserved"
27635
misc_feature /note="Tandem repeat. Forced join. Gap size estimated to
be approximately 320bp by HindIII, SacI and SspI
restriction enzyme digest data."
28011..28380
repeat_region /note="MSTB repeat: matches 1..426 of consensus"
29147..29602
repeat_region /note="LTR24 repeat: matches 1..490 of consensus"
30051..30224
repeat_region /note="2 copies 87 mer 92% conserved"
32247..32977
repeat_region /note="LTPA16 repeat: matches 5423..6154 of consensus"
33453..33754
repeat_region /note="LTR8 repeat: matches 390..691 of consensus"
34281..34449
repeat_region /note="BUR1 repeat: matches 8180..8350 of consensus"
34808..34911
repeat_region /note="4 copies 26 mer 78% conserved"
34811..34902
repeat_region /note="46 copies 2 mer tg 78% conserved"
34813..34902
repeat_region /note="3 copies 30 mer 84% conserved"
34821..34910
repeat_region /note="9 copies 10 mer tgtgtatgta 77% conserved"
34914..34991
repeat_region /note="3 copies 26 mer 79% conserved"
34916..34975
repeat_region /note="6 copies 10 mer tttgtgtgta 78% conserved"
35201..35320
repeat_region /note="MER94 repeat: matches 1..123 of consensus"

```

```

repeat_region 35497..35835
/note="LTPA8 repeat: matches 5819..6162 of consensus"
repeat_region 35836..36095
/note="AluY repeat: matches 39..304 of consensus"
repeat_region 36096..40074
/note="LTPA8 repeat: matches 1897..5819 of consensus"
repeat_region 40075..40337

Query Match 68.0%; Score 17; DB 90; Length 115551;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 aagtcacagaattatt 25
|||||
Db 72877 AAGTCACAGAAATTATT 72861

RESULT 21
AC083945/c
LOCUS
DEFINITION Genomic Sequence For Oryza sativa, Nipponbare Strain, Chromosome X,
AC083945
VERSION
KEYWORDS
SOURCE
ORGANISM
Oryza sativa.
Oryza sativa.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzae;
Oryza.
1 (bases 1 to 147706)
Spiegel, L.A., King, L., Kirchoff, K.A., de la Bastide, M.,
Preston, R.R., Nascimento, L.U., Vil, M.D., Baker, J.P., Miller, B.,
Cunnius, D.M., Kuit, K.H., Rodriguez, S., Santos, L., Zutavern, T.,
Bailja, V.S., Shah, R.S., Bahret, A., Bal, H.P., O'Shaughnessy, A.,
Dedhia, N.N. and McCombie, W.R.
Genomic Sequence For Oryza sativa, Nipponbare Strain, Chromosome X,
Clone OSJNBa0058E19, Complete Sequence
Unpublished
2 (bases 1 to 147706)
McCombie, W.R.
Direct Submission
Submitted (07-OCT-2000) Lita Annenberg Hazen Genome Center, Cold
Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
NY 11724, USA
3 (bases 1 to 147706)
McCombie, W.R.
Direct Submission
Submitted (10-JAN-2001) Lita Annenberg Hazen Genome Center, Cold
Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
NY 11724, USA
4 (bases 1 to 147706)
Bal, H.P., Spiegel, L.A., King, L., Kirchoff, K.A., de la Bastide, M.,
Preston, R.R., Nascimento, L.U., Vil, M.D., Baker, J.P., Miller, B.,
Cunnius, D.M., Kuit, K.H., Rodriguez, S., Santos, L., Zutavern, T.,
Bailja, V.S., Shah, R.S., Bahret, A., O'Shaughnessy, A., Dedhia, N.N.
and McCombie, W.R.
Direct Submission
Submitted (27-FEB-2001) Lita Annenberg Hazen Genome Center, Cold
Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor, NY
11743, USA
Genomic Sequence For Oryza sativa, Nipponbare Strain, Chromosome X,
Clone OSJNBa0058E19, Complete Sequence.
5 (bases 1 to 147706)
Bal, H.P., Spiegel, L.A., King, L., Kirchoff, K.A., de la Bastide, M.,
Preston, R.R., Nascimento, L.U., Vil, M.D., Baker, J.P., Miller, B.,
Cunnius, D.M., Kuit, K.H., Rodriguez, S., Santos, L., Zutavern, T.,
Bailja, V.S., Shah, R.S., Bahret, A., O'Shaughnessy, A., Dedhia, N.N.
and McCombie, W.R.
Direct Submission
Submitted (08-MAR-2001) Lita Annenberg Hazen Genome Center, Cold
Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor, NY
11743, USA

```

REMARK Genomic Sequence For Oryza sativa, Nipponbare Strain, Chromosome X, clone OSJNBa0058E19, Complete Sequence.
On Jan 10, 2001 this sequence version replaced gi:11038452.

COMMENT This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

FEATURES

```

Source
    1..147706
    /organism="Oryza sativa"
    /db_xref="taxon:4530"
    /chromosome="X"
    /clone="OSJNBa0058E19"
    /clone_1lb="HindIII"
    58..5309
    /gene="OSJNBa0058E19.1"
    /note="Histidine kinase-like protein"
    join(58..156,495..1156,1654..2157,2536..2693,2853..2957,
4891..5136,5200..5309)
    /gene="OSJNBa0058E19.1"
    /codon_start=1
    /product="Histidine-kinase like protein"
    /protein_id="AAK13126.1"
    /db_xref="GI:13129468"
    /translation="MAVLVSDVDPVLLIGDPKRFQIITNLVGNMKEFTTEGCHITIRV
HLEEVKRMALDPTSPENITVANSKNTMPYNTLSGLEVANNKRTTLEKFWPDDSS
DAIDVNLVLTVEEDTIGITKDAQRIETFPMAQDGSRTSRGCGIGLSITRKVEID
MGEIETGKPCVSTSTFPTAIEKRNKDPDQIKRCEPTTPDQMGALVDDRCGA
RAVEVYHLRRLGLOCDLAATSESALESNCSSVKSLLMVLVDRKAWMEDDCA
FPRCLDILGTLTKSMQTPKPELLAGSTPPADSDCLLACYSNIRKLPLSYAA
CLSKAGVCLTGRSRDNLVLRVITGKNITLVDDNAVNRVLAAGALKKCAITVCV
DSGEALISRLQPPKFDACFMVQEMQGFETRLVRSVESKINDTIOAGHVSSEIY
GKAHMHVPIILAMTADVIQATEGSSAKNPFDDNNIAKOKSAVNRKQAFDTATYVS
FETTHDSWTFTNTRAGPRGRAPVHAPQVGAIPHPRSLVLAADVOTANMLCN
DTEOKLQTORRSQFLPMRRDCCRTALHDSRSVVVVGCGGTVRGAVTPMWSALDGN
RIG"

repeat_region
4968..7539
    /note="Micropon4 Oryza sativa gene, repeat sequence"
    /product="Micropon4"
    /protein_id="AAK13127.1"
    /db_xref="GI:13129469"
    /translation="MNKLSIALASYLILFLAAAPAAATRTCTFELVTKDGRNMGTD
ARSYLOGPRADKOPGVGDRRRRLREGRRPVGRRGRIHAVALOHGHLRLIM
BOARYVRELDDGAPAGETPVDDAVANGPKNPGFMRSHRSHOPLTTTGWCKSPKY
POLVIMIQIKLPTRAKPTDGTGKPTGSEVLDSSGAAASAAARCTSPISHARRSG
CAPSSPGSSSRRAQDLRVRSKCALFVRAEGASSTTPPEEDTTRAAVAASPL
PALNTNRGLIGQMSPLNLFLKFLFVAPNSFTGEIPQSGFNHMLQITIIYSNNTLQ
RMPNLNANGLVNLGNLNGVIGADLPORFOSLOVNSLNGPIVSQAKITTL
KRPSCLYNDGSIIPMDFAKPIGVYIHLGAKIISGELPSMISGLIPLOKFOELGN
PFYGHITPNLSKLNLDIYSINSTVVRSTICKLKLKLSMLNLELKFPAHSQKD
LDFMNSLANCTELQMSYIGVGGELAPSAAYTSVHREPRVSAKSKNGPSSSPN
PAKKAARSALAPGATLVYTRLPAAAPAEDEGSAFVAYGELDGDGAADDDAAAT
HLAERTPAORSREVRKEKGRGNRKYSTWHPDMGPHGSHADSAA"
7923..9998
    /note="retrosatz 7y3-Gypsy-retroelement"
    18596..19460
    /gene="OSJNBa0058E19.3"
    /note="Hypothetical protein"
    join(18596..18726,18794..18896,18917..18991,19260..19460)
    /gene="OSJNBa0058E19.3"
    /codon_start=1

```

```

gene
    /product="Hypothetical protein"
    /protein_id="AAK13128.1"
    /db_xref="GI:13129470"
    /translation="MNLIVDIRKMSLFLHLHSTSSAANGCLPMSARCRQWTPIG
GHRSSASGSESRQMRSGGCACRALDMMYPKCMCHTHRSSDEVRVNGDSDDAEDVS
ALYHLNSNRNRRDLVAHVHVRGASGAVEAPPNNHCMEALTEIRGDTGIERRMT
REKETIAAQ"
    complement(22790..40376)
    /gene="OSJNBa0058E19.4"
    /note="Putative gag-pol polyprotein"
    complement(22790..23208,24472..24598,24918..25000,
25046..25158,26244..26431,26874..26955,27044..27126,
28396..28467,28767..28900,29535..29647,30186..30251,
30344..30686,31565..31711,32262..32410,33038..33286,
34586..34714,35757..36060,36212..36315,37016..37109,
37580..37812,37836..37921,38120..38188,39131..39252,
39319..39444,39499..40376)
    /gene="OSJNBa0058E19.4"
    /codon_start=1
    /product="Putative gag-pol polyprotein"
    /protein_id="AAK13129.1"
    /db_xref="GI:13129471"
    /translation="MDSIMSGTWVEVRRPYGCKPVGCKWFKRKLPRDGTIEKYAR
LVAGYIQKEGEDEFDTYSPVRLTTRVLALASGLLTHQMDVYTAFLNGLEBE
IYMDOPDGVILESEGWKCLKSLYLKQAPKQMHKRFMDTLTSGAFVNEADKCY
YRGGGEGVILCLVVDILITGSLANVIEEKDPLSEFMTDLRVADVILNIRKLNG
DEGGLTVQSHVYVNSRREYSDCKAAPTPYDPSVLIRKRRIRARQRLYSOTISGL
MYLASATRPDISFAVSKVMRYLKTGMSIGIHAYAGPYLBESTSDSNIISDDELKATS
GSMTEALETALDTATVEAEWMLRELMQDLPVEKVPVALIMNGCARKEAQOTAAAL
SRPYADATRLREELVLEHDSCTSSSSVEIDQNPVLVYGTSCILRNSVRLRCTGRRAI
KELGASASATPVLADYCYCIDAVNMSMEDEVDINGNTGASASSGTFGLNF
REILISGKFPWANSMPPTPKKIDRVLEFEEVKKVYVPMERKKRAGDPAPALI
EHKGFKFTVYMANRILALVAKYTRISQTAFFRGNINMEALIIHETLHEMCKKSD
GVLILKLYEKSSTYDENMFELOALRMSSTYTFWRWIDQVSGSVAAKPYCIGDPLV
LGPGTLTVLEKSTLLIKRGVETKITYILREKAPAVCAASAASAPSLPRQPSA
AAAPATVAPAGHRRGAVATVLLPSSVASDCSSAALIDDPKTAASSTSLPSSKSP
FESLCLILACFERREGDYDPGVAKECELSYGLVLTQMTQCSAMAHDDVYTCG
VYSGCYSGPTLRKRCGODASVEASRQETVLTTRLCRCPAPMLAHOVEKHPRAFR
OVGSHRAHARPCHTSDEPTISDGADERGVITRLTKGPIQDQVMPASDKNADKT
RLRVDCLAMPPLVLPDGLRTNNTNIGVSAIIRKSTGRTDGTQDVRTIACVRRRAAR
HIPSNDGSPFERMDKRRPTANAAGCAASSGVPNRRKQARPASSPACQAGQRLVLP
LAGYGARPGSTNRPFKLOEISSRPKVAKLSPSPRPNRORRRESGKRLRSLAQTWERP
SLGTPGATVGNMGKGVPTYPHILMERESRYAFANVASCALNRSMTGRTGTQDVRT
IACVRRRAARHIPSNDGERSIYGRNLEAEVAVNVALTLDQDREFKQKNKG
VYPRKNTFMKLIKPLKISFLMYLAKERYIEFGLTMAKRWPLLKQVDEMYKSKCLL
EKRALFEFSMTVVTMBPSTIEAFHRAPSMASRIVVTTIKOHDEAGCEKVKFEEDS
MOCCGRRTKRFASKRLSLIOSSORRKGFLWPTPTTSCDEHMSAEVAVNGRCFAAVY
LGRAGEPPVGEMMAVMAKRRIDAYCSIGVGRG"
25681..29265
    /note="retro02 retroelement"
    30888..31407
    /note="RIR84 Oryza sativa gypsy-type retrotransposon"
    complement(30963..32572)
    /note="RIR8 Oryza sativa gypsy-type retrotransposon"
    31529..31692
    /note="truncator transposon from Oryza sativa"
    complement(31804..32083)
    /note="RIR8 Oryza sativa gypsy-type retrotransposon"
    32562..34237
    /note="truncator transposon"
    complement(32638..33296)
    /note="RIR8 Oryza sativa gypsy-type retrotransposon"
    complement(34475..43713)
    /gene="OSJNBa0058E19.5"
    /note="Putative gag-pol polyprotein"
    complement(join(34475..34662,35774..35971,37596..37812,
39105..39266,39507..40550,40983..42461,42516..42653,
42714..43024,43070..43269,43327..43469,43702..43713))
    /gene="OSJNBa0058E19.5"
    /codon_start=1
    /product="Putative gag-pol polyprotein"
    /protein_id="AAK13130.1"
    /db_xref="GI:13129472"

```


FEATURES	source
*	161460 163515: contig of 2056 bp in length
*	163516 163615: gap of 100 bp
*	163616 166614: contig of 2999 bp in length
*	166615 166714: gap of 100 bp
*	166715 169056: contig of 2342 bp in length
*	169057 169156: gap of 100 bp
*	169157 171527: contig of 2371 bp in length
*	171528 171627: gap of 100 bp
*	171628 173248: contig of 1621 bp in length
*	173249 173348: gap of 100 bp
*	173349 174815: contig of 1467 bp in length
	Location/Qualifiers
	1..174815

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q23"
/clone="RP11-635b11"
1..58823
misc_feature
/note="assembly_fragment"
58924..73243
/note="assembly_fragment"
73344..83572
/note="assembly_fragment"
83673..94547
/note="assembly_fragment"
94648..104597
/note="assembly_fragment"
104698..114347
/note="assembly_fragment"
114448..122080
/note="assembly_fragment"
122181..128465
/note="assembly_fragment"
128566..135791
/note="assembly_fragment"
135892..142934
/note="assembly_fragment"
143035..147999
/note="assembly_fragment"
148100..151734
/note="assembly_fragment"
151835..155576
/note="assembly_fragment"
155677..159193
/note="assembly_fragment"
159294..161359
/note="assembly_fragment"
161460..163515
/note="assembly_fragment"
163616..166614
/note="assembly_fragment"
166715..169056
/note="assembly_fragment"
169157..171527
/note="assembly_fragment"
171628..173248
/note="assembly_fragment"
173349..174815
/note="assembly_fragment"
/note="assembly_fragment"
BASE COUNT 52310 a 30923 c 32183 g 57399 t 2000 others
ORIGIN
Query Match 68.0%; Score 17; DB 82; Length 174815;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
Oy 7 gcaagtcacagaattat 23
DB 44384 GCAAGTCACAGATTAT 44368

```

```

RESULT 24
AC002523/C 228516 bp DNA PRI 05-MAY-1998
LOCUS Homo sapiens Xq28 BACs 360 F12, GSHB-555C13, complete sequence.
DEFINITION AC002523 AC002374 AC002373
AC002523.1 GI:2815518
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS Muzny,D., Ansari-Lari,M.A., Timms,K.M., Yu,W., Dugan,S., Lu,J.,
Shen,Y., Howland,K., Liu,W., Perez,L., Ding,Y., Haywood,M.,
Jain,A., Leal,B., Logan,O., Nguyen,V., Savage,L., Shen,H.,
Worley,K., Chen,E., Forcum,J., Aronson,A.D., Gorrell,J.H.,
Brundage,E., Di,W., Chinault,C., Nelson,D. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 228516)
Chiu,M.W.
REFERENCE Direct Submission
JOURNAL Submitted (10-SEP-1997) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 228516)
Worley,K.C.
REFERENCE Direct Submission
JOURNAL Submitted (28-JAN-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 228516)
Worley,K.C.
REFERENCE Direct Submission
JOURNAL Submitted (05-MAY-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
Sequencing is completed to a minimum standard of double strand
coverage with a minimum of 2 clones and 2 reads with no ambiguities
or 2 chemistries with a minimum of 2 clones and 3 reads with no
ambiguities. If the sequence quality does not meet this standard,
it will be indicated in the annotation.
COMMENT
The repeat regions shown were identified using RepeatMasker by
Adrian Smiit.
Sequence similarities were identified using Powerblast by Jinghui
Zhang.
Exon/Intron boundaries of identified genes were chosen if there
were canonical splice junctions that maintained sequence continuity
across the splice junctions.
Beginning of sequence overlaps with AF007262, end of sequence
overlaps with AF011889.
FEATURES
Source
Location/Qualifiers
1..228516
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="360 F12, GSHB-555C13"
/chromosome="X"
1..149
/notice="Positions 104645 to 104793 in AF007262 overlap with
positions 1-149 in this clone"
/note="Region: Overlap with AF007262"
repeat_region complement(487..547)
/rpt_family="MER5B"
repeat_region complement(1270..1577)
/rpt_family="MLT1E"
repeat_region complement(1653..1708)
/rpt_family="MLT1D"
repeat_region complement(2282..2306)
/rpt_family="AT_rich"
repeat_region 3316..3617
/rpt_family="MLT1B"
repeat_region 3618..3651

```

```

/rpt_family="(CA)n"
complement(3658..3944)
/rpt_family="AluX"
3949..4012
/rpt_family="MLT1B"
complement(4098..4386)
/rpt_family="LINE2"
4723..4784
/rpt_family="MIR"
5630..5727
/rpt_family="(GA)n"
7337..7584
/rpt_family="LINE2"
7678..7709
/rpt_family="(CA)n"
7710..7863
/rpt_family="LINE2"
complement(7981..8151)
/rpt_family="LINE2"
complement(8177..8274)
/rpt_family="LINE2"
complement(10118..10192)
/rpt_family="LINE2"
complement(10400..10443)
/rpt_family="AT_rich"
complement(10668..10695)
/rpt_family="AT_rich"
11811..11896
/rpt_family="LINE2"
12526..12605
/rpt_family="LINE2"
14892..14932
/rpt_family="MLT1A2"
15361..15661
/rpt_family="AluX"
16706..16765
/rpt_family="AT_rich"
17193..17764
/rpt_family="L1PA7"
17811..18912
/rpt_family="L1PA7"
complement(18968..19018)
/rpt_family="L1"
19034..19085
/rpt_family="L1"
complement(19084..19727)
/rpt_family="L1"
19737..19971
/rpt_family="L1"
19971..20406
/rpt_family="LINE3A"
20452..20500
/rpt_family="AT_rich"
20899..21008
/rpt_family="(CAAA)n"
21061..21098
/rpt_family="(TA)n"
complement(21283..21343)
/rpt_family="AT_rich"
complement(21527..22106)
/rpt_family="MLT1F"
complement(22286..22326)
/rpt_family="AT_rich"
23171..23316
/rpt_family="MER44A"
23311..23626
/rpt_family="MER44C"
complement(23681..23757)
/rpt_family="AT_rich"
24613..24924
/rpt_family="AluX"
complement(24927..25278)
/rpt_family="MSTA"

```



```

/map="21921.2"
/clone="R132H24, 5' partial"
/clone_11b="R1C1-11 BAC library"
complement(1162. .1526)
/repeat_region
  /rpt_family="LINE/L1"
  /rpt_type=DISPERSED
  /note="L1PA13"
  1527. .1893
/repeat_region
  /note="L1PA13"
  /rpt_family="LINE/L1"
  /rpt_type=DISPERSED
  1905. .2287
/repeat_region
  /note="L2"
  /rpt_family="LINE/L2"
  /rpt_type=DISPERSED
  2386. .2549
/repeat_region
  /note="L2"
  /rpt_family="LINE/L2"
  /rpt_type=DISPERSED
  complement(4287. .4410)
  /note="L2"
  /rpt_family="LINE/L2"
  /rpt_type=DISPERSED
  complement(4489. .4670)
  /note="L2"
  /rpt_family="LINE/L2"
  /rpt_type=DISPERSED
  complement(6498. .6745)
  /note="L1MA9"
  /rpt_family="LINE/L1"
  /rpt_type=DISPERSED
  7600. .7882
/repeat_region
  /note="AluSx"
  /rpt_family="SINE/Alu"
  /rpt_type=DISPERSED
  complement(7912. .8041)
  /note="AluDb"
  /rpt_family="SINE/Alu"
  /rpt_type=DISPERSED
  complement(8745. .9054)
  /note="AluSx"
  /rpt_family="SINE/Alu"
  /rpt_type=DISPERSED
  complement(9504. .9794)
  /note="AluY"
  /rpt_family="SINE/Alu"
  /rpt_type=DISPERSED
  9983. .10564
/repeat_region
  /note="L1MB8"
  /rpt_family="LINE/L1"
  /rpt_type=DISPERSED
  10720. .10787
  /note="(TAGA)n"
  /rpt_family="Simple_repeat"
  /rpt_type=TANDEM
  11137. .11436
/repeat_region
  /note="AluY"
  /rpt_family="SINE/Alu"
  /rpt_type=DISPERSED
  13135. .13345
  /standard_name="D21S1914"
  /note="AFM344W65"
  Accession No. 251475"
  13259. .13313
  /note="(TG)n"
  /rpt_family="Simple_repeat"
  /rpt_type=TANDEM
  complement(14329. .14632)
  /note="L1MA3"
  /rpt_family="LINE/L1"
  /rpt_type=DISPERSED
  14625. .14852
  /note="L1MA3"

```

```

/repeat_region
  /rpt_family="LINE/L1"
  /rpt_type=DISPERSED
  14853. .15154
  /note="AluY"
  /rpt_family="SINE/Alu"
  /rpt_type=DISPERSED
  15155. .15634
  /note="L1MA3"
  /rpt_family="LINE/L1"
  /rpt_type=DISPERSED
  complement(15694. .15808)
  /note="L1MA6"
  /rpt_family="LINE/L1"
  /rpt_type=DISPERSED
  15840. .16145
  /note="AluY"
  /rpt_family="SINE/Alu"
  /rpt_type=DISPERSED
  17627. .18220
  /note="L1"
  /rpt_family="LINE/L1"
  /rpt_type=DISPERSED
  18201. .18826
  /note="L1"
  /rpt_family="LINE/L1"
  /rpt_type=DISPERSED
  18827. .19126
  /note="AluSx"
  /rpt_family="SINE/Alu"
  /rpt_type=DISPERSED
  19127. .19352
  /note="L1"
  /rpt_family="LINE/L1"
  /rpt_type=DISPERSED
  19353. .20555
  /note="L1PA7"
  /rpt_family="LINE/L1"
  /rpt_type=DISPERSED
  20557. .21024
  /note="L1P5"
  /rpt_family="LINE/L1"
  /rpt_type=DISPERSED
  21026. .21055
  /note="(TG)n"
  /rpt_family="Simple_repeat"
  /rpt_type=TANDEM
  complement(21414. .21879)
  /note="L1MA5A"
  /rpt_family="LINE/L1"

```

```

Query Match      68.0%; Score 17; DB 91; Length 340000;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      9 aagtcacagaattattt 25
          |||
Db 141812 AAGTCACAGAAATATT 141796

```

```

RESULT 26
HSAAPAF20/c      HSAAPAF20      605 bp      DNA      PRI      03-AUG-2000
LOCUS
DEFINITION
ACCESSION
AF098889
VERSION
AF098889.1 GI:4191520
KEYWORDS
SEGMENT
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 605)
Bata,S., Oliver,H., Renault,B., Montgomery,K., Dutta,S., Rao,P.,

```

Houldsworth, J., Kucherlapati, R., Wang, X., Chaganti, R.S. and Murty, V.V.
 Genetic analysis of the APAF1 gene in male germ cell tumors
 Genes Chromosomes Cancer 28 (3), 258-268 (2000)
 MEDLINE 20320642
 PUBMED 10862031
 REFERENCE 2 (bases 1 to 605)
 AUTHORS Bala, S., Oliver, H., Renault, B., Dutta, S., Montgomery, K., Kucherlapati, R., Wang, X., Chaganti, R.S.K. and Murty, V.V.
 TITLE Direct Submission
 JOURNAL Submitted (14-OCT-1998) Pathology, College of Physicians & Surgeons of Columbia University, 630 West 168th Street, New York, New York 10032, USA

FEATURES
 source Location/Qualifiers
 1..605
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 <1..70
 /gene="APAF1"
 /number=12
 71..>605
 /gene="APAF1"
 /number=12
 Intron
 BASE COUNT 184 a 93 c 134 g 192 t 2 others
 ORIGIN

Query Match 64.0%; Score 16; DB 93; Length 605;
 Best Local Similarity 100.0%; Pred. NO. 19;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 9 aagtcacagaattat 24
 ||||||||||||||||
 Db 425 AAGTCACAGATTATT 410

RESULT 27
 LOCUS G07181 1829 bp DNA STS 19-OCT-1995
 DEFINITION human STS WI-9211.
 ACCESSION G07181
 VERSION G07181.1 GI:860426
 KEYWORDS STS sequence; primer; sequence tagged site.
 SOURCE human STS derived from sequences in dbEST and the Unigene collection.
 ORGANISM Homo sapiens
 Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 1829)
 AUTHORS Hudson, T.
 TITLE Whitehead Institute/MIT Center for Genome Research; Physically Mapped STS
 JOURNAL Unpublished (1995)
 COMMENT

Contact: Thomas Hudson
 Whitehead Institute/MIT Center for Genome Research
 Whitehead Institute for Biomedical Research
 9 Cambridge Center, Cambridge MA 02142 USA
 Tel: 617 252 1900
 Fax: 617 252 1902
 Email: thudson@genome.wi.mit.edu
 Primer A: CAGCCACACTAATGATTAACG
 Primer B: GACAGAAATGATGTATTAATGCG
 STS size: 111
 PCR profile:
 Presoak:
 Denaturation: 56 degrees C
 Annealing: 56 degrees C
 Polymerization:
 PCR Cycles: 35

Thermal Cycler:
 Protocol:
 Template: 10 ng
 Primer: each 5 pm
 dNTPs: each 4 mM
 Tag Polymerase: 0.025 units/ul
 Total Vol: 20 ul

Buffer:
 MgCl2: 1.5 mM
 KCl: 50 mM
 Tris-HCL: 10 mM
 pH: 9.3

Prepared with primer pairs derived from D13644 -- Unigene.
 FEATURES
 source Location/Qualifiers
 1..1829
 /organism="Homo sapiens"
 STS
 primer_bind 891..1001
 primer_bind 891..912
 primer_bind complement(977..1001)
 BASE COUNT 539 a 321 c 320 g 590 t 59 others
 ORIGIN

Query Match 64.0%; Score 16; DB 54; Length 1829;
 Best Local Similarity 100.0%; Pred. NO. 17;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 8 caagtcacagaattat 23
 ||||||||||||||||
 Db 1689 CAAGTCACAGATTATT 1674

RESULT 28
 LOCUS D13644 4602 bp mRNA PRI 10-JUL-1997
 DEFINITION Human mRNA for KIAA0019 gene, complete cds.
 ACCESSION D13644
 VERSION D13644.1 GI:1531551
 KEYWORDS KIAA0019.
 SOURCE Homo sapiens male myeloblast cell_line KG-1 cDNA to mRNA.
 ORGANISM Homo sapiens
 Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 4602)
 AUTHORS Nomura, N.
 JOURNAL Direct Submission
 Submitted (11-NOV-1992) to the DDBJ/EMBL/GenBank databases. Nobuo Yana, Kisarazu, Chiba 292, Japan (E-mail:cdna@fokeazusa.or.jp), URL: http://www.kazusa.or.jp, Tel: 0438-52-3930, Fax: 0438-52-3931)
 REFERENCE 2 (bases 1 to 4602)
 AUTHORS Nomura, N., Miyajima, N., Kawarabayashi, Y. and Tabata, S.
 TITLE Prediction of new human genes by entire sequencing of randomly sampled cDNA clones
 JOURNAL Unpublished (1994)
 REFERENCE 3 (sites)
 AUTHORS Nomura, N., Miyajima, N., Sazuka, T., Tanaka, A., Kawarabayashi, Y., Sato, S., Nagase, T., Seki, N., Ishikawa, K. and Tabata, S.
 TITLE Prediction of the coding sequences of unidentified human genes. I. The coding sequences of 40 new genes (K1A0001-K1A0040) deduced by analysis of randomly sampled cDNA clones from human immature myeloid cell line KG-1
 JOURNAL DNA Res. 1 (1), 27-35 (1994)
 MEDLINE 96051387
 REMARK Erratum: [[published erratum appears in DNA Res 1995 Aug 31:2(4):211]]

REFERENCE 4 (sites)
 AUTHORS Nomura, N., Miyajima, N., Sazuka, T., Tanaka, A., Kawarabayashi, Y., Sato, S., Nagase, T., Seki, N., Ishikawa, K. and Tabata, S.
 TITLE Prediction of the coding sequences of unidentified human genes. I. The coding sequences of 40 new genes (K1A0001-K1A0040) deduced by

```

analysis of randomly sampled cDNA clones from human immature
myeloid cell line KG-1 (supplement)
JOURNAL MEDLINE 96051389
REFERENCE 5 (bases 1 to 4602)
AUTHORS Matsukawa,B., Wong,W.-T., Seki,N., Nagase,T., Nomura,N.,
Robbins,K.C., and Di Fiore,P.P.
TITLE RN-tre identifies a family of tre-related proteins displaying a
novel potential protein binding domain
JOURNAL MEDLINE 96293402
COMMENT On Sep 10, 1996 this sequence version replaced gi:286016.
FEATURES
    source
        1..4602
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /cell_line="KG-1"
        /cell_type="myoblast"
        /chromosome="10"
        /map="10p13"
        /sex="male"
        <1..279
        280..2766
        /gene="K1A0019"
        280..2766
        /gene="K1A0019"
        /codon_start=1
        /product="protein related N-terminus of tre oncogene"
        /protein_id="BA02807.1"
        /db_xref="GI:2104571"
        /translation="MNSDQVALKLAERAEIVAKYDGRGAEIEPMEDADYLVYV
        TDRGFLHEELPDHNAVEROKHLEIERTKMKLMKGEYKNTKEPHRIYKGIP
        LQLRGEVALLLETPKMEETRDLYSKLHRAGCSPDIDIDVNRTRFDHIMFD
        RYGVKOOSLFHVLAAVSYNTEVGYCGMSOITLLMNMEDAPALVKKLTSQSPH
        AMHGFVGGPPKLLRPOEHKRLINKPKSKLKHLDQSEITTSYTTKMFQOCLDPT
        PPTLNRIWDIYIEGGERVLTAMSTYTIKLMKLMKMELEVEFQETLAKDFFE
        DDFVLEQIISMTLKRKLDLPEPKDEDEYKPKPLQQLPELOSQGVHLSNGQSV
        GRPSPLAGRSRSGAPHRHSHSPHPOSRTGPRAOPPRKKSVEESKLIKLEADQ
        RKLPSGDDSSROYNHAANONSATNITKEFVKNKPSDVATERTAKYTWEKG
        RAHPLALAVYVGPAAEVSVNVRKMKALDAEDGRSTASOVNVPGPSARV
        EALRRASQSRHALYPPSPRKHAPSSSPSKVKNKTFEYVQPPSHARVQSOLDGAR
        GLAHPSPSLVYPPSSRIEVLVDGTAGGIGSGSGSPKNGKLLTPVDYLPNRTWSEVS
        KSYRPSPLVYPPSSRIEVLVDGTAGGIGSGSGSPKNGKLLTPVDYLPNRTWSEVS
        YTYRPTGOSWTDASRGNLPKTSFQLAFQDHGLPAVSVSPVRYKASPAEDAS
        PSQYPSGPPPAHYRNRLDLSIOESVLL"
        2767..4602
BASE COUNT 1340 a 1041 c 999 g 1222 t
ORIGIN
Query Match 64.0%; Score 16; DB 91; Length 4602;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 8 caagtcacagaattat 23
|||||
Db 4462 CAAAGTCACAGATTAT 4447

RESULT 29
LOCUS LLA012388
DEFINITION Lactococcus lactis dar gene and genes encoding five hypothetical
proteins.
ACCESSION AJ012388
VERSION AJ012388.1 GI:6165401
KEYWORDS acetoin(diacetyl)reductase; dar gene; ORF1; ORF3; ORF4; ORF5; ORF6;
putative.
SOURCE Lactococcus lactis.
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Lactococcus.
REFERENCE 1 (bases 1 to 5469)

AUTHORS Aungpraphanornchai,P., Griffin,H.G. and Gasson,M.J.
TITLE Cloning, DNA sequence analysis, and deletion of a gene encoding
diacetyl-acetoin reductase from Lactococcus lactis
JOURNAL MEDLINE 20112344
REFERENCE 2 (bases 1 to 5469)
AUTHORS Aungpraphanornchai,P.
TITLE Direct Submission
JOURNAL Submitted (28-Oct-1998) Aungpraphanornchai P., Genetics and
Microbiology, Institute of Food Research, Colney lane, Norwich,
Norfolk, NR4 7UA, UK
FEATURES
    source
        1..5469
        /organism="Lactococcus lactis"
        /sub_species="cremoris"
        /strain="WGI363"
        /db_xref="taxon:1358"
        77..80
        /note="ORF1"
        87..950
        /note="ORF1"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="CAB59825.1"
        /db_xref="GI:6165402"
        /translation="MNPKNRITIAVAVLILVALVAFPSLSNQGCVKASAGETVKG
        INSGDKQDEVMKSVANTAKERTYDLKLVFVSDYNQPEALLSDIDINAFQSYNV
        KTWNKARKSDIYAVVNTYITPMHITSKLSLSDKESTVAIPDASNESALPLVQ
        SAGLLKLTTSDSKLGPLDITENPHOLEFKKVDASQTPRALDSVALSVNNVNTAA
        SLPSKSVMEPLNKTSAQYINFIATTSKKNKYYKEVAKAYASKATEKAIKEQPD
        GGLPLPMDLKL"
        1005..1009
        /gene="dar"
        1005..1789
        /gene="dar"
        1019..1789
        /gene="dar"
        /codon_start=1
        /transl_table=11
        /product="diacetyl (acetoin) reductase"
        /protein_id="CAB59826.1"
        /db_xref="GI:6165403"
        /translation="MGLKLAIVTSGSGIGEGIVYRLAKDGEFAVAADINPQAKNVAK
        TLIDKGEFAKAVLVDAANSAGFOLIEQAVSPGFLAVFVNNAGVAFISDIIDSPSR
        VERLIPVYNKGTGYWIOAAAGVQKRGGRGRIINAASLASVSGASASASRAIR
        GLEQSAKELAKDIYTVNAVYNGIYRTAMREDIDKRTQKIQKISTEQQQNCLEETIL
        GREGREDVAEVYVAFASDALEYITGQSLVDGMRFN"
        1794..1798
        /note="ORF3"
        1805..2665
        /note="ORF3"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="CAB59827.1"
        /db_xref="GI:6165404"
        /translation="MNVKNRRILITIIIVFIIIVGIGFAFSHGSKNSKVSXKIVKIG
        LMPGKQEDVIMKVOYOKNAKDQGITLEFVNTPDDEDEPNKALVNHVEVDLFOHAYL
        KSWKNANGNINVISIGDITITPHIYSTKGVDEIPDKSTAIAPNDITNESRALYVVK
        NAGLIKUDTSRGVATYKDIRNPRLSLIKELIDSQTPRALDSVAANVINYNPALISAK
        NSDKRSTOEPLNEDSAQWINFITANQSDKNKYYKEVAKAYEONKINADILIKKEYPD
        GELPAMNLIK"
        2776..2781
        /note="ORF4"
        2789..3895
        /note="ORF4"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="CAB59828.1"
        /db_xref="GI:6165405"
        /translation="MTATIELNLSVQPHQKGRGLVTAVKNAATLIEKGDITGVIGYSG

```

DB	3204	AAGTCACAGATTATT	3219	Query Match Best Local Similarity 64.0%; Score 16; DB 3; Length 5469; Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	9	aagtcacagattatt	24	
RESULT	30			
LOCUS	AL161800/c			
DEFINITION	AL161800	43741 bp	DNA	PRI 04-APR-2001
ACCESSION	AL161800	Human DNA sequence from clone RP11-318C17 on chromosome 20. Contains part of a novel gene and GSSs, complete sequence.		
KEYWORDS	AL161800.10	GI:8246908		
SOURCE	HTG.			
ORGANISM	human.			
REFERENCE	Homo sapiens			
TITLE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	Wilson, S.			
JOURNAL	Direct Submission Submitted (15-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk			
COMMENT	On Jun 4, 2000 this sequence version replaced gi:8217673. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping above. only a small overlap as described above.			

FEATURES	Source	Location/Qualifiers
repeat_region	repeat_region	1..43741
repeat_region	repeat_region	/organism="Homo sapiens"
repeat_region	repeat_region	/db_xref="taxon:9606"
repeat_region	repeat_region	/chromosome="20"
repeat_region	repeat_region	/clone="RP11-318C17"
repeat_region	repeat_region	/clone_lib="RPCT-11.2"
repeat_region	repeat_region	3..293
repeat_region	repeat_region	/note="AluX repeat: matches 1..291 of consensus"
repeat_region	repeat_region	390..683
repeat_region	repeat_region	/note="AluSg repeat: matches 1..286 of consensus"
repeat_region	repeat_region	788..3327
repeat_region	repeat_region	/note="L1 repeat: matches 2348..2710 of consensus"
repeat_region	repeat_region	3411..3820
repeat_region	repeat_region	/note="match: GSS: Em:AQ371040"
repeat_region	repeat_region	4157..4477
repeat_region	repeat_region	/note="L2 repeat: matches 2348..2710 of consensus"
repeat_region	repeat_region	complement(6523..6967)
repeat_region	repeat_region	/note="match: GSS: Em:AQ236127"
repeat_region	repeat_region	complement(6607..6950)
repeat_region	repeat_region	/note="match: GSS: Em:AQ110431"
repeat_region	repeat_region	6676..6820
repeat_region	repeat_region	/note="MIR repeat: matches 79..229 of consensus"
repeat_region	repeat_region	7231..7405
repeat_region	repeat_region	/note="MIR repeat: matches 42..206 of consensus"
repeat_region	repeat_region	7416..7551
repeat_region	repeat_region	/note="L2 repeat: matches 2584..2710 of consensus"
repeat_region	repeat_region	9521..9605
repeat_region	repeat_region	/note="MIR repeat: matches 433..523 of consensus"
repeat_region	repeat_region	complement(join(9542..9883,13385..>13593))
repeat_region	repeat_region	/gene="DA318C17.1"
repeat_region	repeat_region	/note="continued from d3974N19.2 in Em:ALJ38808
repeat_region	repeat_region	match: ESTs: Em:AL542749 Em:AL569592"
repeat_region	repeat_region	/evidence=not_experimental
repeat_region	repeat_region	/product="DA318C17.1 (novel transcript)"
repeat_region	repeat_region	complement(9542)
repeat_region	repeat_region	/gene="DA318C17.1"
repeat_region	repeat_region	complement(join(9542..9883,13385..13593))
repeat_region	repeat_region	/gene="DA318C17.1"
repeat_region	repeat_region	complement(9565..9570)
repeat_region	repeat_region	/gene="DA318C17.1"
repeat_region	repeat_region	9605..9884
repeat_region	repeat_region	/note="MIR repeat: matches 59..137 of consensus"
repeat_region	repeat_region	9885..10320
repeat_region	repeat_region	/note="MER4B repeat: matches 1..437 of consensus"
repeat_region	repeat_region	10325..10397
repeat_region	repeat_region	/note="MER4C repeat: matches 393..465 of consensus"
repeat_region	repeat_region	10682..11085

```

repeat_region /note="L2 repeat: matches 1981. .2406 of consensus"
11167. .11657
repeat_region /note="L2 repeat: matches 2263. .2750 of consensus"
11872. .12002
repeat_region /note="L2 repeat: matches 2621. .2750 of consensus"
12681. .12976
repeat_region /note="AluYb repeat: matches 1. .305 of consensus"
14480. .15774
repeat_region /note="L2 repeat: matches 1424. .2749 of consensus"
15775. .16067
repeat_region /note="AluXg repeat: matches 1. .310 of consensus"
16068. .16216
repeat_region /note="L2 repeat: matches 1285. .1424 of consensus"
16713. .16865
repeat_region /note="L2 repeat: matches 2545. .2699 of consensus"
17484. .17764
repeat_region /note="AluXg repeat: matches 6. .296 of consensus"
17813. .18013
repeat_region /note="L1MB2 repeat: matches 5966. .6164 of consensus"
18014. .18267
repeat_region /note="AluSc repeat: matches 1. .252 of consensus"
18268. .18773
repeat_region /note="L1MB2 repeat: matches 5450. .5966 of consensus"
19067. .19374
repeat_region /note="AluXg1 repeat: matches 1. .304 of consensus"
19513. .19811
repeat_region /note="AluYbA8 repeat: matches 1. .305 of consensus"
21304. .21613
repeat_region /note="AluSg repeat: matches 1. .308 of consensus"
22398. .23343
repeat_region /note="MT1A1 repeat: matches 1. .353 of consensus"
23736. .23865
repeat_region /note="AluSg repeat: matches 1. .130 of consensus"
24570. .25025
repeat_region /note="MT1C repeat: matches 1. .466 of consensus"
25235. .26166
repeat_region /note="L1M4 repeat: matches 3073. .4010 of consensus"
complement(26577. .27248)
misc_feature /note="match: GSS: Em:AQ036701"
28315. .28629
repeat_region /note="AluSg repeat: matches 1. .313 of consensus"
28933. .29096
repeat_region /note="MT1I repeat: matches 245. .409 of consensus"
29893. .30110
repeat_region /note="MIR repeat: matches 25. .245 of consensus"
complement(30032. .30494)
misc_feature /note="match: GSS: Em:AQ23457"
31501. .31807
repeat_region /note="AluXg repeat: matches 1. .307 of consensus"
32165. .32267
repeat_region /note="L2 repeat: matches 2554. .2646 of consensus"
32268. .32752
repeat_region /note="MER6B repeat: matches 1. .486 of consensus"
32753. .33089
repeat_region /note="L2 repeat: matches 2213. .2554 of consensus"
33351. .33651
repeat_region /note="AluXg repeat: matches 1. .304 of consensus"
33662. .33966
repeat_region /note="AluY repeat: matches 1. .295 of consensus"
34730. .34818
repeat_region /note="MIR repeat: matches 43. .145 of consensus"
complement(34862. .35287)
misc_feature /note="match: GSS: Em:AQ040224"
complement(35239. .35778)
misc_feature /note="match: GSS: Em:AQ634531"
complement(35374. .35798)
misc_feature /note="match: GSS: Em:AQ565906"
36173. .36533
repeat_region /note="match: GSS: Em:AQ231637"
36392. .36553
repeat_region /note="MER5A repeat: matches 4. .175 of consensus"
36668. .36961
repeat_region /note="AluSg repeat: matches 1. .297 of consensus"

```

```

repeat_region 37033. .37258
/note="MIR repeat: matches 15. .252 of consensus"
misc_feature 38080. .38507
/note="match: GSS: Em:B36377"
misc_feature 38087. .38612
/note="match: GSS: Em:AQ680468"
repeat_region 38349. .38418
/note="35 copies 2 mer tg 74% conserved"
repeat_region 38694. .38979
/note="AluJo repeat: matches 1. .286 of consensus"
misc_feature 40625. .41129
/note="match: GSS: Em:AQ680145"
repeat_region 41704. .41819
/note="MIR repeat: matches 57. .177 of consensus"
repeat_region 41842. .42034
/note="L1MC4 repeat: matches 7779. .7969 of consensus"
repeat_region 42037. .42226
/note="MT1I repeat: matches 45. .247 of consensus"
misc_feature complement(42848. .43207)
/note="match: GSS: Em:AQ385536"
repeat_region 42966. .43258
/note="L2 repeat: matches 1138. .1493 of consensus"
repeat_region 43259. .43555
/note="AluYb repeat: matches 10. .303 of consensus"
repeat_region 43556. .43741
/note="L2 repeat: matches 970. .1138 of consensus"
BASE COUNT 12960 a 8117 c 8797 g 13867 t
ORIGIN

```

```

Query Match 64.0%; Score 16; DB 90; Length 43741;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 aagtcacagaattatt 24
|||||
Db 10623 AAGTCACAGATTATT 10608

```

```

RESULT 31
AC074073/c 68894 bp DNA HTG 13-JUL-2000
LOCUS Homo sapiens chromosome 3 clone RP11-79116 map 3, LOW-PASS SEQUENCE
DEFINITION
AC074073
ACCESSION AC074073
VERSION AC074073.1 GI:9170900
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 68894)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 3, clone RP11-79116
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 68894)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G.,
Campiano,A., Castle,A., Choquet,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domingo,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glende,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howard,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Lahocque,K., Lamazares,R., Landers,T., Lehotzky,J.,
Levine,R., Liu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McQuirk,A., McKernan,K., McPheters,R.,
Meldrum,T., Menous,L., Minova,T., Miranda,C., Mlenka,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rotman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,

```

TITLE
JOURNAL
COMMENT

Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Vassiliev,S., Theodore,J., Titrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Submitted (13-JUL-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/repeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L7127

Center clone name: 791_L16

* NOTE: This record contains 83 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1 747: contig of 747 bp in length
* 748 847: gap of 100 bp
* 848 1579: contig of 732 bp in length
* 1580 1679: gap of 100 bp
* 1680 2425: contig of 746 bp in length
* 2426 2525: gap of 100 bp
* 2526 3254: contig of 729 bp in length
* 3255 3354: gap of 100 bp
* 3355 4075: contig of 721 bp in length
* 4076 4175: gap of 100 bp
* 4176 4892: contig of 717 bp in length
* 4893 4992: gap of 100 bp
* 4993 5717: contig of 725 bp in length
* 5718 5817: gap of 100 bp
* 5818 6558: contig of 741 bp in length
* 6559 6658: gap of 100 bp
* 6659 7371: contig of 713 bp in length
* 7372 7471: gap of 100 bp
* 7472 8209: contig of 738 bp in length
* 8210 8309: gap of 100 bp
* 8310 9046: contig of 737 bp in length
* 9047 9146: gap of 100 bp
* 9147 9888: contig of 742 bp in length
* 9889 9988: gap of 100 bp
* 9989 10723: contig of 735 bp in length
* 10724 10823: gap of 100 bp
* 10824 11553: contig of 730 bp in length
* 11554 11653: gap of 100 bp
* 11654 12391: contig of 738 bp in length
* 12392 12491: gap of 100 bp
* 12492 13214: contig of 723 bp in length
* 13215 13314: gap of 100 bp
* 13315 14034: contig of 720 bp in length
* 14035 14134: gap of 100 bp
* 14135 14664: contig of 730 bp in length
* 14665 14964: gap of 100 bp
* 14965 15700: contig of 736 bp in length
* 15701 15800: gap of 100 bp
* 15801 16535: contig of 735 bp in length
* 16536 16635: gap of 100 bp
* 16636 17357: contig of 722 bp in length
* 17358 17457: gap of 100 bp
* 17458 18192: contig of 735 bp in length

* 18193 18292: gap of 100 bp
* 18293 19030: contig of 738 bp in length
* 19031 19130: gap of 100 bp
* 19131 19857: contig of 727 bp in length
* 19858 19957: gap of 100 bp
* 19958 20677: contig of 720 bp in length
* 20678 20777: gap of 100 bp
* 20778 21514: contig of 737 bp in length
* 21515 21614: gap of 100 bp
* 21615 22336: contig of 722 bp in length
* 22337 22436: gap of 100 bp
* 22437 23190: contig of 754 bp in length
* 23191 23290: gap of 100 bp
* 23291 24024: contig of 734 bp in length
* 24025 24124: gap of 100 bp
* 24125 24859: contig of 735 bp in length
* 24860 24959: gap of 100 bp
* 24960 25700: contig of 741 bp in length
* 25701 25800: gap of 100 bp
* 25801 26544: contig of 744 bp in length
* 26545 26644: gap of 100 bp
* 26645 27367: contig of 723 bp in length
* 27368 27467: gap of 100 bp
* 27468 28196: contig of 729 bp in length
* 28197 28296: gap of 100 bp
* 28297 29028: contig of 732 bp in length
* 29029 29128: gap of 100 bp
* 29129 29843: contig of 715 bp in length
* 29844 29943: gap of 100 bp
* 29944 30633: contig of 690 bp in length
* 30634 30733: gap of 100 bp
* 30734 31472: contig of 739 bp in length
* 31473 31572: gap of 100 bp
* 31573 32319: contig of 747 bp in length
* 32320 32419: gap of 100 bp
* 32420 33151: contig of 732 bp in length
* 33152 33251: gap of 100 bp
* 33252 33977: contig of 726 bp in length
* 33978 34077: gap of 100 bp
* 34078 34818: contig of 741 bp in length
* 34819 34918: gap of 100 bp
* 34919 35632: contig of 714 bp in length
* 35633 35732: gap of 100 bp
* 35733 36474: contig of 742 bp in length
* 36475 36574: gap of 100 bp
* 36575 37300: contig of 726 bp in length
* 37301 37400: gap of 100 bp
* 37401 38137: contig of 737 bp in length
* 38138 38237: gap of 100 bp
* 38238 38960: contig of 723 bp in length
* 38961 39060: gap of 100 bp
* 39061 39789: contig of 729 bp in length
* 39790 39889: gap of 100 bp
* 39890 40617: contig of 728 bp in length
* 40618 40717: gap of 100 bp
* 40718 41442: contig of 725 bp in length
* 41443 41542: gap of 100 bp
* 41543 42278: contig of 736 bp in length
* 42279 42378: gap of 100 bp
* 42379 43117: contig of 739 bp in length
* 43118 43217: gap of 100 bp
* 43218 43961: contig of 744 bp in length
* 43962 44061: gap of 100 bp
* 44062 44786: contig of 725 bp in length
* 44787 44886: gap of 100 bp
* 44887 45614: contig of 728 bp in length
* 45615 45714: gap of 100 bp
* 45715 46458: contig of 744 bp in length
* 46459 46558: gap of 100 bp
* 46559 47292: contig of 734 bp in length
* 47293 47392: gap of 100 bp
* 47393 48107: contig of 715 bp in length
* 48108 48207: gap of 100 bp

```

* 48208 48935: contig of 728 bp in length
* 48936 49035: gap of 100 bp
* 49036 49765: contig of 730 bp in length
* 49766 49865: gap of 100 bp
* 49866 50604: contig of 739 bp in length
* 50605 50704: gap of 100 bp
* 50705 51434: contig of 730 bp in length
* 51435 51534: gap of 100 bp
* 51535 52266: contig of 732 bp in length
* 52267 52366: gap of 100 bp
* 52367 53096: contig of 730 bp in length
* 53097 53196: gap of 100 bp
* 53197 53936: contig of 740 bp in length
* 53937 54036: gap of 100 bp
* 54037 54770: contig of 734 bp in length
* 54771 54870: gap of 100 bp
* 54871 55574: contig of 704 bp in length
* 55575 55674: gap of 100 bp
* 55675 56392: contig of 718 bp in length
* 56393 56492: gap of 100 bp
* 56493 57227: contig of 735 bp in length
* 57228 57327: gap of 100 bp

```

Query Match Best Local Similarity 100.0%; Score 16; DB 75; Length 68894;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 agtcacagaattatt 25
Db 58360 AGTCACAGATTATT 58345

```

RESULT 32
AC012679/c AC012679 105937 bp DNA PLN 19-JAN-2001
DEFINITION Arabidopsis thaliana chromosome 1 BAC F25P22 genomic sequence,
AC012679 complete sequence.
AC012679 AC012679.5 GI:12324196
HTG
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosid II;
Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 105937)
Lin, X., Kaul, S., Town, C.D., Benito, M., Creasy, T.H., Haas, B.J.,
Wu, D., Matli, R., Konning, C.M., Koo, H., Fujii, C.Y., Utterback, T.R.,
Barnstead, M.E., Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.
Arabidopsis thaliana chromosome 1 BAC F25P22 genomic sequence
Unpublished
2 (bases 1 to 105937)
Lin, X. and Kaul, S.
Direct Submission
Submitted (03-NOV-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
3 (bases 1 to 105937)
Town, C.D. and Kaul, S.
Direct Submission
Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
On Jan 19, 2001 this sequence version replaced gl:12280804.
Address all correspondence to: atetigr.org

```

COMMENT
BAC clone F25P22 is from Arabidopsis thaliana chromosome 1
The orientation of the sequence is from Spe to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan (Chris Burge,
http://CCR-081.mt.edu/GENSCAN.html), GenemarkHMM (Mark Borodovsky,
http://genemark.biology.gatech.edu/Genemark/), Glimmer4 (a variant
of Glimmer3, see Mhaela Perlea,
http://www.tigr.org/softlab/glimmer4/glimmer4.htm/glimmer4.html, and

FEATURES

GenesPlicer (Mhaela Perlea and Steven Salzberg, contact
mperlea@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted as
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
Simple repeats are identified by RepeatMasker (Arian Smith,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).
Location/Qualifiers
1..105937

```

source
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/map="ngal11"
/clone="F25P22"
join(<1..76,158..299,430..710,789..950,1031..1084,
1171..1512)
/gene="F25P22.1"
1..1512
/gene="F25P22.1"
/note="similar to
S-adenosyl-methionine-sterol-C-methyltransferase
GB:AC34851 [Nicotiana tabacum]"
join(<1..76,158..299,430..710,789..950,1031..1084,
1171..1344)
/gene="F25P22.1"
/codon_start=2
/product="putative
S-adenosyl-methionine-sterol-C-methyltransferase, 5'
partial: 1-1344"
/protein_id="AA052075.1"
/db_xref="GI:12324209"
/translation="NSTELSVSCKICGAYVRNKKNONOICLMQKVSDDRGQRF
LDWVQYKSSGILKRYERFGEVSTGLETTEFEVDMLDKPGQVLDVGCIGGGD
YVAENFDVAVVGDIDLSVNMISFALHAIQICSVFEVADCTKKEYPNDTPVVISRD
TIIHIDKPEALFEREYKWKLPQGVKVLITDYCSPTSPDFAIYIKKRGYDLHDVQAY
GMLRDAPGEALFEVVAEDRTDQPMKVLKREDAVEKEKEEFISPSKEDYEDDILGKMSK
LRSSSGEOKWGLFTAKRN"
join(<6321..6579,6690..6820,6910..7146,7259..7502,
7508..>7751)
/gene="F25P22.2"
6321..7751
/gene="F25P22.2"
/note="F25P22.2"
/note="contains Pfam profile: PF00657 lipase/acylhydrolase
with GDSL-like motif"
join(6321..6579,6690..6820,6910..7146,7259..7502,
7508..7751)
/gene="F25P22.2"
/codon_start=1
/product="putative lipase/acylhydrolase; 6321-7751"
/protein_id="AA052082.1"
/db_xref="GI:12324216"
/translation="MNCIMFEKMLAFSIFLYVGNQOSTGNSVALFAFGDSIL
DTGNMNLISVSKVNEYYPGRDFIGRATGRGNGRVSDDIAEGIGKNIIPAYRD
YLMNNDLTGVCFSAGSGIDPTARTGSIIVSQVDVFOYITRLNGVGNQOAN
AYISNVAIVYISAGNNDIATTFPTGARRLQVYLPAVNOIVSWFMDIKSLYDMGARK
FVAMGTIPGICPGARATIRACELFYVNGAANFNQNSADINDNLCATRPQAKYVYDM
YNPLGLIINPDASGFIADVADACCTPTHILPCLDASRVEVDVAHPQKSTETIAPQ
IIEIKAKLA"
join(<9376..9496,9643..9967,10550..>10898)
/gene="F25P22.3"
9376..10898
/gene="F25P22.3"
/note="contains Pfam profile: PF00314 thaumatin family"
join(9376..9496,9643..9967,10550..10898)
/gene="F25P22.3"

```



```

/codon_start=1
/product="thumatin-like protein; 9376-10898"
/protein_id="AAG52086.1"
/db_xref="GI:12324220"
/translation="MTNEKEVALPELKLSEKSPMASSSLFLPLLULLLVASASTV
ITYNKTYVWPGLIPSSGOSLLAGGFKLSNRAVTTIQLPLMVGREMGRCGCFDR
SGRGATGDCGGSLCNGAGVPPATLAEITLIGDMDFVSLVDGYLANSTIPVK
GTGKTCAGACVSDLNMCPVGLQVRSRDGTQVACKSACSAFNSPRCCTGTFGNPOS
CKPTASKIFKVAQPKAYSAYDDPTSTATCSKANVVTFCPRGR"
<12692..>13183
/gene="F25P22.4"
12692..13183
/gene="F25P22.4"
/notes="similar to calmodulin GB:P02596 [Renilla
reniformis]; contains Pfam profile: PF00036 EF hand (4
copies)"
12692..13183
/gene="F25P22.4"
/codon_start=1
/product="putative calmodulin; 12692-13183"
/protein_id="AAG52088.1"
/db_xref="GI:12324222"
/translation="MANTNLESTNKSTPTSTMEIKKVFDRKDANGDKISVSLGNV
FKSMGTSTYEEELNRYLDEIDIDCDGFIHQEFATICRSSSAVEITREAFDLYDNKN
GLISSSEIHKVIRLQMTCSVEDCYRMIGHVDTDGNVNEFEFOKMMSSPELVKGTV
ANS"
join(<14977..15197,15451..>15931)
/gene="F25P22.5"
14977..15931
/gene="F25P22.5"
/notes="contains Pfam profile: PF00071 ras family"
join(14977..15197,15451..15931)
/gene="F25P22.5"
/codon_start=1
/product="putative ras-related GTP-binding protein;
14977-15931"
/protein_id="AAG52089.1"
/db_xref="GI:12324223"
/translation="MAEDTTEECYLYKFAVLIGSAVGSNLSRSKDEFRDPSK
TIGVEFARNVHGDKLIKAQIMDTAGCGREFATISSYRGALGALLIYDITRTTFD
NIRKKLFELDRANETVYLVGNKSDLSKSEVEDESKTAEBSGLFELTSLVEN
VVEEAFVIMGRHEVYQRIASEKNSGAATPHNGNGCTVLPVGRKEIVNIHEVT
ATQPLSSSSNCCFK"
complement(join(<16797..16967,17064..17297,17388..17469,
17546..17667,17706..17773,17866..17935,18448..18772,
19179..19223,19427..>19725))
/gene="F25P22.7"
complement(16797..19725)
/gene="F25P22.7"
complement(16797..19725)
complement(join(16797..16967,17064..17297,17388..17469,
17546..17667,17706..17773,17866..17935,18448..18772,
19179..19223,19427..19725))
/gene="F25P22.7"
/codon_start=1
/product="unknown protein; 19725-16797"
/protein_id="AAG52066.1"
/db_xref="GI:12324200"
/translation="MATLFTATVPSSHHRVSPSQPKOSLSOSISVFTTEPOTAY
VTLQEQLDWITSPVTRRFGAGFTWAGLAFVGSQMKSRIDVFOEDNTRGL
EKQOEITLPNGIRYDLOVGSATPSSGLVVEDVGVGHGNEVDFDGGKSLA
AMYDSRPYSGKCGCQIEHLRSMKAGKRKVIIPSLGCGDNVEGCOLEFPSPAT
DYILIEYOFIEFVITLAFKPOVDVFASTNVILAVITLVKSMHROVLVLNV
VVMGLGIEFILMLITONGEDRDEDRGNITRIITFMVLQVVMVTVSLPTLVNAS
DGGSLKPADVIGTWMVYVGLLEAADDQKLSFKNSPKNRCKWCDVGYMKSRIPNT
FEBMLMWGIEFAASPVLEGAELVTFGLFTLLLVSGILPLEVPCSLPLVDN
P"
complement(join(<20191..20329,20451..20627,20721..20770,
20870..20938,21048..21146,21239..21317,21582..21682,
21776..21844,21957..21997,22081..23243,23382..23485,
23568..24008,24102..>24662))
/gene="F25P22.8"
complement(20191..24662)
/gene="F25P22.8"

```

```

/notes="contains Pfam profile: PF00069 eukaryotic protein
kinase domain"
complement(join(20191..20329,20451..20627,20721..20770,
20870..20938,21048..21146,21239..21317,21582..21682,
21776..21844,21957..21997,22081..23243,23382..23485,
23568..24008,24102..>24662))
/gene="F25P22.8"
/codon_start=1
/product="putative protein kinase; 24662-20191"
/protein_id="AAG52069.1"
/db_xref="GI:12324203"
/translation="MKVKEETLKNLGDGVLRPYDHCSSITWSMKMKKFLKLIHSP
NQSDAEKISITFKSNHHSIDVSSSSPSRSHSPKIFSGLSNMLSSVGRKTP
SPNSFNKRNRAATVDTVYVNGSGEYHDLGSKDPAVEENOILALELSAREDEATQ
IEATKQFSLGSCAPENSFAELIAYRYNNYNTLGVDKILIDGFYDLGVLANSSAERTP

```

Query Match 64.0%; Score 16; DB 12; Length 105937;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 actcaagaattatt 25
Db 55362 actcaagaattatt 55347

RESULT 33
ABO20878
LOCUS ABO20878 110525 bp DNA PRI 21-MAY-1999
DEFINITION Homo sapiens genomic DNA of 9q32 anti-oncogene of flat epithelium
cancer, segment 10/10.
ACCESSION ABO20878
VERSION ABO20878.1 GI:4003398
KEYWORDS
SOURCE
ORGANISM Homo sapiens DNA.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS Nakamura,Y., Isomura,M. and Daigo,Y.
TITLE Homo sapiens 1,010,525bp genomic DNA of 9q32 anti-oncogene of flat
epithelium cancer region
JOURNAL Published Only in Database (1998) In press
REFERENCE 2 (bases 1 to 110525)
AUTHORS Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
TITLE Direct Submission
JOURNAL Submitted (30-NOV-1998) to the DDBJ/EMBL/GenBank databases. Miki
Hirakawa, Japan Science and Technology Corporation (JST), Advanced
Database Department; 5-3, Yonban-cho, Chiyoda-ku, Tokyo 102-0028,
Japan (E-mail:mikae@tokyo.jst.go.jp, Tel:81-3-5214-8491,
Fax:81-3-5214-8470)
COMMENT This sequence is conducted by Japanese Foundation for Cancer
Research as a JST seauencing team
Principal Investigator: Yusuke Nakamura Ph.D.
Phone:+81-3-5449-5372, Fax:+81-3-5449-5433,
Yusuke@jgc.iins.u-tokyo.ac.jp
The sequence is submitted by:Human Genome Sequencing in ALIS
project of JST
Japan Science and Technology Corporation (JST)
5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081 Japan
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site
(http://www-alis.tokyo.jst.go.jp/HGS/top.html)
or send email to webmaster@www-alis.tokyo.jst.go.jp.
Location/Qualifiers
1..110525
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="9q32"
80338..80503
/standard_name="D9S1126"
/notes="CHLC.GCAT11B01.P3342;The location is between each

FEATURES
source
STS

```

flanking site of PCR primers."
/db_xref="GDB:686181"
STS
86433..86587
/Note="STS6871;The location is between each flanking site
of PCR primers."
/db_xref="GDB:4569679"
104545..104691
/Note="SHGC-18832;The location is between each flanking
site of PCR primers."
/db_xref="GDB:6458849"
BASE COUNT 32237 a 22817 c 23211 g 32215 t 45 others
ORIGIN
Query Match 64.0%; Score 16; DB 85; Length 110525;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0:
OY 9 aagtcacagattatt 24
|||||
Db 39879 AACTCACAGATTATT 39894

RESULT 34
AC011248 111312 bp DNA PRI 28-APR-2000
LOCUS Homo sapiens 12 BAC RP11-14P21 (Roswell Park Cancer Institute Human
DEFINITION BAC library) complete sequence.
AC011248
AC011248 GI:7658299
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 111312)
Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
Bodola,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
Holloway,C., Horak,H., Issar,A., Jackson,L.E., Jackson,L., Jia,Y.,
Jones,M., Kelly,S., Kneitz,S., Kondejewski,N., Kong,Y., Kovar,C.,
Lau,S., Leal,B., Lee,E., Li,Z., Licharge,O., Liu,J., Liu W.,
Logan,O., Lozano,R.J., Lu,J., Luchter,R., Marondel,I., Martin,R.,
Martinez,C., McLeod,M.P., Mel,G., Merscher,S., Miller,A.,
Montgomery,K.T., Morgan,M., Morris,S., Nash,S., Nelson,A.,
Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G., Parish,B., Paxton,S.,
Payton,B., Perez,L., Pu,L.L., Quiles,M., Reiter,D., Rives,M.,
Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H., Shim,C., Simon,M.,
Sparks,A., Stamps,A., Sucgang,R., Tabor,P., Taylor,T., Vasquez,L.,
Vinson,R., Vo,Q., Wahbah,M., Watlington,S., Weinstein,G.,
Weinstock,I.R., Williamson,A., Worley,K., Wren,J., Wrensford,G.,
Xiang,A.M., Yang,R., Yu,W., Zhou,X., Kucherlapati,R., Nelson,D., and
Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 111312)
Worley,K.C.
Direct Submission
Submitted (05-OCT-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 111312)
Worley,K.C.
Direct Submission
Submitted (28-APR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Apr 28, 2000 this sequence version replaced gi:6996901.
COMMENT
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email

```

gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

```

----- Summary Statistics -----
Contig Length: 150286
Phrap values in estimate: 149476
Average error rate (BCM-Phrap estimate): 0.000180006
Fraction of Phrap values less than 40 : 0.0291418
Number of consensus changing edits: 15
Number of N's in consensus : 1

```

```

----- Consensus Changing Edits -----
Position Original+Context Edited+Context
22441 aaagagaaga(n)gagggctcca aaagagaaga(a)gagggctcca
72231 aaagagcca(n)antttant aaagagcca(a)ttttatag
72233 taagccana(n)nttanc aaagccata(c)ttttaagtc
72234 aaagccana(n)tttanc aaagccata(c)ttttaagtc
72239 caantttta(n)ttcttgta caattttta(a)ttcttgta
79241 naantttta(n)ctcttgta latittta(a)gctcttgta
79245 ttantnct(n)gtgaataa tttaatgc(c)gtgaataa
79279 aaagctgta(n)gctgaataa aaagctgta(a)gctgaataa
79440 agcctggcca(n)catgtaaa agcctggcca(a)catgtaaa
79920 tctcaaaa(n)caaaagaag tctcaaaa(a)caaaagaag
94117 gtttccag(n)tcctggcaga gtttccag(t)tcctggcaga
95790 gcgaagctc(n)tcctcaaaa gcgaagctc(a)tcctcaaaa
111017 ttaagttcc(n)tcagctgaac ttaagttcc(a)tcagctgaac
115887 atgcata(n)caacaacaa atgcata(a)caacaacaa
129846 ggaagctgaa(n)cgagtgatc ggaagctgaa(g)cgagtgatc

```

```

----- Distribution of Quality < 40 Bases -----
10001 *
9001 *
8001 *
7001 *

```



```

FEATURES      *      97971 98070: gap of 100 bp
                *      98071 118636: contig of 20566 bp in length.
                Location/Qualifiers
                source      .118636
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="6"
                        /clone_1b="RP3-356B20"
                        /clone_1b="RPC1-3"
                        1..24283
                        /note="assembly_fragment:01399
                        fragment_chain:1
                        clone_end:17
                        vector_side:left"
                misc_feature      24384..26993
                        /note="assembly_fragment:00194
                        fragment_chain:1"
                misc_feature      27094..56307
                        /note="assembly_fragment:00731
                        fragment_chain:1"
                misc_feature      56408..59297
                        /note="assembly_fragment:00116
                        fragment_chain:1"
                misc_feature      59398..80206
                        /note="assembly_fragment:00023
                        fragment_chain:1"
                misc_feature      80307..97970
                        /note="assembly_fragment:00217
                        fragment_chain:1"
                misc_feature      98071..118636
                        /note="assembly_fragment:00177
                        fragment_chain:1
                        clone_end:SP6
                        vector_side:right"

BASE COUNT      36924 a 21288 c 22673 g 37150 t 601 others
ORIGIN
Query Match      64.0%: Score 16; DB 82; Length 118636;
Best Local Similarity 100.0%: Pred. No. 13;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 agtcacagaattattt 25
|||||
Db 81402 AGTCACAGATTATT 81417

RESULT 36
AL159168 129010 bp DNA PRI 30-SEP-2000
LOCUS      Human DNA sequence from clone RP11-401H23 on chromosome 9, complete
DEFINITION
ACCESSION      AL159168
VERSION      AL159168.15 GI:10443391
KEYWORDS      HTG.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 129010)
AUTHORS      Kimberley, A.
TITLE      Direct Submission
JOURNAL      Submitted (26-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT      Requests: clonerequests@sanger.ac.uk
On Oct 1, 2000 this sequence version replaced gi:10178449.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria

```

```

FEATURES      Source
                Location/Qualifiers
                1..129010
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="9"
                /clone_1b="RP11-401H23"
                /clone_1b="RPC1-11.2"
                39045 a 27222 c 26534 g 36089 t

BASE COUNT      39045 a 27222 c 26534 g 36089 t
ORIGIN
Query Match      64.0%: Score 16; DB 90; Length 129010;
Best Local Similarity 100.0%: Pred. No. 13;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 agtcacagaattattt 25
|||||
Db 102456 AGTCACAGATTATT 102471

RESULT 37
AC019289 139218 bp DNA HTG 08-APR-2000
LOCUS      Homo sapiens chromosome 4 clone RP11-203119 map 4, WORKING DRAFT
DEFINITION
ACCESSION      AC019289
VERSION      AC019289.3 GI:7528059
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 139218)
AUTHORS      Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beckert, R., Bida, F.,
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,
Chen, Y., Chiswick, M., Collins, S., Collins, A., Cooke, P.,
DeRubeis, K., Dewar, K., Dominko, M., Doyle, M., Feneberg, J.,
Fitzgerald, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardner, S., Grant, G., Hargis, B., Heaford, A., Horton, L.,
Howland, J.C., Johnson, R., Jones, C., Kann, L., Karas, A., Klein, J.,
Lander, T., Lebeck, J., Levine, R., Liu, C., Liu, G., Locke, K.,
McDonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
McPheters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J.,

```

as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; SW, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/Hbp/Chr9>

RP11-401H23 is from the library RPc1-11.2 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/vectors/PBACe3.6>

IMPORTANT: This sequence is not the entire insert of clone RP11-401H23. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP11-401H23 is at 1 in this sequence. The true left end of clone RP11-16121 is at 128911 in this sequence. The true right end of clone RP11-38608 is at 19333 in this sequence.

TITLE
JOURNAL
COMMENT

Norman, C.H., O'Connor, T., O'Donnell, P., Olivari, T.M., Peterson, K.,
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rotman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Tittell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
Zimmer, A. and Zody, M.

Submitted (31-DEC-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 8, 2000 this sequence version replaced gi:6778536.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: 14306

Center clone name: 203.L19

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 13255 bases at least Q40

Consensus quality: 135869 bases at least Q30

Consensus quality: 137053 bases at least Q20

Insert size: 140000; agarose-fp

Insert size: 138118; sum-of-contigs

Quality coverage: 4.9 in Q20 bases; agarose-fp

Quality coverage: 5.0 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1      879: contig of 879 bp in length
*      880 979: gap of 100 bp
*      980 2319: contig of 1340 bp in length
*      2320 2419: gap of 100 bp
*      2420 6972: contig of 4553 bp in length
*      6973 7072: gap of 100 bp
*      7073 11233: contig of 4161 bp in length
*      11234 11333: gap of 100 bp
*      11334 18294: contig of 6961 bp in length
*      18295 18394: gap of 100 bp
*      18395 24919: contig of 6525 bp in length
*      24920 25019: gap of 100 bp
*      25020 32971: contig of 7952 bp in length
*      32972 33071: gap of 100 bp
*      33072 44845: contig of 11774 bp in length
*      44846 44945: gap of 100 bp
*      44946 60683: contig of 15738 bp in length
*      60684 60783: gap of 100 bp
*      60784 77274: contig of 16491 bp in length
*      77275 77374: gap of 100 bp
*      77375 102352: contig of 24978 bp in length
*      102353 102452: gap of 100 bp
*      102453 139218: contig of 36766 bp in length.

```

FEATURES

source

```

1. 139218
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-203119"
/clone.lib="RP11-11 Human Male BAC"
misc_feature
1. 879

```

```

/note="assembly_fragment
clone_end:SP6
vector_side:left"
980..2319
/note="assembly_fragment
clone_end:T7
vector_side:right"
2420..6972
/note="assembly_fragment"
7073..11233
/note="assembly_fragment"
11334..18294
/note="assembly_fragment"
18395..24919
/note="assembly_fragment"
25020..32971
/note="assembly_fragment"
33072..44845
/note="assembly_fragment"
44946..60683
/note="assembly_fragment"
60784..77274
/note="assembly_fragment"
77375..102352
/note="assembly_fragment"
102453..139218
/note="assembly_fragment"
25775 c 25817 g 43205 t 1102 others
BASE COUNT 43319 a 25775 c 25817 g 43205 t 1102 others
ORIGIN

```

```

Query Match 64.0% Score 16; DB 65; Length 139218;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 10 agtcacagaattatt 25
|||||
Db 103832 AGTCACAGAAATTATT 103817

```

```

RESULT 38
AC090105/c DNA 11-MAR-2001
LOCUS Homo sapiens chromosome 8 clone RP11-629E1 map 8, *** SEQUENCING IN
DEFINITION PROGRESS ***, 51 unordered pieces.
AC090105
VERSION AC090105.2 GI:13273432
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 152452)
AUTHORS Biren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 8, clone RP11-629E1
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 152452)
AUTHORS Biren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Boguslavsky, L., Bouhgalter, B., Brown, A.,
Cammarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,
Collins, A., Cooke, P., DeArrelano, K., Dewar, K., Diaz, J.S.,
Dodgson, S., Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,
Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, M., Iliev, I., Johnson, R.,
Jones, C., Karatas, A., Larocque, K., Lamazares, R., Landers, T.,
Lehoczky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P.,
Margulis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
McPheters, R., Melnick, J., Meneus, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S., Severy, P.,

```



```

misc_feature      3157..3978
                    /note="assembly_fragment"
misc_feature      4079..4745
                    /note="assembly_fragment"
misc_feature      4846..6069
                    /note="assembly_fragment"
misc_feature      6170..7193
                    /note="assembly_fragment"
misc_feature      7294..8578
                    /note="assembly_fragment"
misc_feature      8679..9808
                    /note="assembly_fragment"
misc_feature      9909..11353
                    /note="assembly_fragment"
misc_feature      11454..12575
                    /note="assembly_fragment"
misc_feature      12676..14311
                    /note="assembly_fragment"
misc_feature      14412..15641
                    /note="assembly_fragment"
misc_feature      15742..16712
                    /note="assembly_fragment"
misc_feature      16813..18001
                    /note="assembly_fragment"

Query Match      64.0%; Score 16; DB 77; Length 152452;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      10 agtcacagaattattt 25
        |||||||
Db      4588 AGTCACAGATTATT 4573

RESULT 39
AC067998/c      AC067998      153221 bp      DNA      HTG      14-JUN-2000
LOCUS      Homo sapiens chromosome 3 clone RP11-627D21 map 3, WORKING DRAFT
DEFINITION
SEQUENCE 29 unordered pieces.
AC067998
AC067998.2      GI:8516111
KEYWORDS
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 153221)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguski,M., Bouckgeater,B., Brown,A., Burkett,G.,
Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., DeCarliano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehocck,J.,
Levine,R., Liu,C., Liu,G., Locke,K., MacDonald,P., Margulis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,
Meldrum,J., Menus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Strange-Rhmann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Teste,S., Theodore,J., Tirelli,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
JOURNAL
Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome

```

COMMENT

```

Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 14, 2000 this sequence version replaced gi:7656755.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8575
Center clone name: 627.D.21
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 137105 bases at least Q40
Consensus quality: 145055 bases at least Q30
Consensus quality: 148395 bases at least Q20
Insert size: 157000; agarose-1p
Insert size: 150421; sum-of-contrigs
Quality coverage: 3.9 in Q20 bases; agarose-1p
Quality coverage: 4.1 in Q20 bases; sum-of-contrigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1      1623: contrig of 1623 bp in length
*
*      1624 1723: gap of 100 bp
*
*      1724      3380: contrig of 1657 bp in length
*
*      3381 3480: gap of 100 bp
*
*      3481      5348: contrig of 1868 bp in length
*
*      5349 5448: gap of 100 bp
*
*      5449      7335: contrig of 1887 bp in length
*
*      7336 7435: gap of 100 bp
*
*      7436      9535: contrig of 2100 bp in length
*
*      9536 9635: gap of 100 bp
*
*      9636 12801: contrig of 3166 bp in length
*
*      12802 12901: gap of 100 bp
*
*      12902      15358: contrig of 2457 bp in length
*
*      15359 15458: gap of 100 bp
*
*      15459      17864: contrig of 2406 bp in length
*
*      17865 17964: gap of 100 bp
*
*      17965      19498: contrig of 1534 bp in length
*
*      19499 19598: gap of 100 bp
*
*      19599      21802: contrig of 2204 bp in length
*
*      21803 21902: gap of 100 bp
*
*      21903      24102: contrig of 2200 bp in length
*
*      24103 24202: gap of 100 bp
*
*      24203      26673: contrig of 2471 bp in length
*
*      26674 26773: gap of 100 bp
*
*      26774      30361: contrig of 3588 bp in length
*
*      30362 30461: gap of 100 bp
*
*      30462      34817: contrig of 4356 bp in length
*
*      34818 34917: gap of 100 bp
*
*      34918      38602: contrig of 3685 bp in length
*
*      38603 38702: gap of 100 bp
*
*      38703      43613: contrig of 4911 bp in length
*
*      43614 43713: gap of 100 bp
*
*      43714      48515: contrig of 4802 bp in length
*
*      48516 48615: gap of 100 bp
*
*      48616      53390: contrig of 4775 bp in length
*
*      53391 53490: gap of 100 bp
*
*      53491      58254: contrig of 4764 bp in length
*
*      58255 58354: gap of 100 bp
*
*      58355      65768: contrig of 7414 bp in length
*
*      65769 65868: gap of 100 bp

```

```

* 65869 70737: contig of 4869 bp in length
* 70738 70837: gap of 100 bp
* 70838 77128: contig of 6291 bp in length
* 77129 77228: gap of 100 bp
* 77229 84888: contig of 7660 bp in length
* 84889 84988: gap of 100 bp
* 84989 94687: contig of 9699 bp in length
* 94688 94787: gap of 100 bp
* 94788 105941: contig of 11154 bp in length
* 105942 106041: gap of 100 bp
* 106042 113657: contig of 7616 bp in length
* 113658 113757: gap of 100 bp
* 113758 122759: contig of 9002 bp in length
* 122760 122859: gap of 100 bp
* 122860 136138: contig of 13279 bp in length
* 136139 136238: gap of 100 bp
* 136239 153221: contig of 16983 bp in length.

```

FEATURES

```

Source
1..153221
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="3"
  /map="3"
  /clone_1lb="RP11-627D21"
  1..1623
    misc_feature
      /note="assembly_fragment"
      1724..3380
        misc_feature
          /note="assembly_fragment"
          3481..5348
            misc_feature
              /note="assembly_fragment"
              5449..7335
                misc_feature
                  /note="assembly_fragment"
                  7436..9535
                    misc_feature
                      /note="assembly_fragment"
                      9636..12801
                        misc_feature
                          /note="assembly_fragment"
                          12902..15358
                            misc_feature
                              /note="assembly_fragment"
                              15459..17864
                                misc_feature
                                  /note="assembly_fragment"
                                  17965..19498
                                    misc_feature
                                      /note="assembly_fragment"
                                      19599..21802
                                        misc_feature
                                          /note="assembly_fragment"
                                          21903..24102
                                            misc_feature
                                              /note="assembly_fragment"
                                              24203..26673
                                                misc_feature
                                                  /note="assembly_fragment"
                                                  26774..30361
                                                    misc_feature
                                                      /note="assembly_fragment"
                                                      30462..34817
                                                        misc_feature
                                                          /note="assembly_fragment"
                                                          34918..38602
                                                            misc_feature
                                                              /note="assembly_fragment"
                                                              38703..43613
                                                                misc_feature
                                                                  /note="assembly_fragment"
                                                                  43714..48515
                                                                    misc_feature
                                                                      /note="assembly_fragment"
                                                                      48616..53390
                                                                        misc_feature
                                                                          /note="assembly_fragment"
                                                                          53491..58254
                                                                            misc_feature
                                                                              /note="assembly_fragment"
                                                                              58355..65768
                                                                                misc_feature
                                                                                  /note="assembly_fragment"
                                                                                  65869..70737
                                                                                    misc_feature
                                                                                                                                 /note="assembly_fragment
                                                                                                                                 clone_end:SP6
                                                                                                                                 vector_side:left"
                                                                                                                                 70838..77128
                                                                                                                                 /note="assembly_fragment"
                                                                                                                                 77229..84888
                                                                                                                                 /note="assembly_fragment"

```

```

misc_feature 84989..94687
              /note="assembly_fragment"
misc_feature 94788..105941
              /note="assembly_fragment"
              clone_end:TT7
              vector_side:right"
misc_feature 106042..113657
              /note="assembly_fragment"
misc_feature 113758..122759
              /note="assembly_fragment"
misc_feature 122860..136138
              /note="assembly_fragment"
              /note="assembly_fragment"
              136239..153221
              /note="assembly_fragment"
BASE COUNT 42805 a 32811 c 32675 g 42124 t 2806 others
ORIGIN
Query Match 64.0%; Score 16; DB 73; Length 153221;
Best Local Similarity 100.0%; Pred.No. 12;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 10 agtcacagaattattt 25
Db 14609 AGTCACAGATTATT 14594

```

```

RESULT 40
AC019258/C AC019258 155211 bp DNA HTG 13-JUL-2000
LOCUS Homo sapiens clone RP11-2M23, LOW-PASS SEQUENCE SAMPLING.
AC019258
AC019258.2 GI:7209488
VERSION HTG; HTGS_PHASE0.
KEYWORDS SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 155211)
Birten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barina,N., Beckerley,J., Bida,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepe,I., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dekrellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczyk,J., Levine,R., Lien,C., Liu,G., Locke,K.,
Macdonald,P., Margulis,N., McEwan,P., McGuck,A., McKernan,K.,
McPheters,R., Meldrum,J., Meneses,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
2 (bases 1 to 155211)
Birten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barina,N., Beckerley,J., Bida,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepe,I., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dekrellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczyk,J., Levine,R., Lien,C., Liu,G., Locke,K.,
Macdonald,P., Margulis,N., McEwan,P., McGuck,A., McKernan,K.,
McPheters,R., Meldrum,J., Meneses,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.

```

```

TITLE Direct Submission
JOURNAL Submitted (31-DEC-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 8, 2000 this sequence version replaced gi:5649456.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

```

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2672

```


Center clone name: 2_M_23

* NOTE: This record contains 177 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
757 856: gap of 100 bp in length
857 1619: contig of 763 bp in length
1620 1719: gap of 100 bp
1720 2498: contig of 779 bp in length
2499 2598: gap of 100 bp
2599 3355: contig of 757 bp in length
3356 3455: gap of 100 bp
3456 4201: contig of 746 bp in length
4202 4301: gap of 100 bp
4302 5053: contig of 752 bp in length
5054 5153: gap of 100 bp
5154 5910: contig of 757 bp in length
5911 6010: gap of 100 bp
6011 6779: contig of 769 bp in length
6780 6879: gap of 100 bp
6880 7629: contig of 750 bp in length
7630 7729: gap of 100 bp
7730 8485: contig of 756 bp in length
8486 8585: gap of 100 bp
8586 9348: contig of 763 bp in length
9349 9448: gap of 100 bp
9449 10225: contig of 777 bp in length
10226 10325: gap of 100 bp
10326 11083: contig of 758 bp in length
11084 11183: gap of 100 bp
11184 11964: contig of 781 bp in length
11965 12064: gap of 100 bp
12065 12830: contig of 766 bp in length
12831 12930: gap of 100 bp
12931 13692: contig of 762 bp in length
13693 13792: gap of 100 bp
13793 14557: contig of 765 bp in length
14558 14657: gap of 100 bp
14658 15412: contig of 755 bp in length
15413 15512: gap of 100 bp
15513 16282: contig of 770 bp in length
16283 16382: gap of 100 bp
16383 17131: contig of 749 bp in length
17132 17231: gap of 100 bp
17232 17980: contig of 749 bp in length
17981 18080: gap of 100 bp
18081 18839: contig of 759 bp in length
18840 18939: gap of 100 bp
18940 19701: contig of 762 bp in length
19702 19801: gap of 100 bp
19802 20565: contig of 764 bp in length
20566 20665: gap of 100 bp
20666 21415: contig of 750 bp in length
21416 21515: gap of 100 bp
21516 22286: contig of 771 bp in length
22287 22386: gap of 100 bp
22387 23161: contig of 775 bp in length
23162 23261: gap of 100 bp
23262 24022: contig of 761 bp in length
24023 24122: gap of 100 bp
24123 24872: contig of 750 bp in length
24873 24972: gap of 100 bp
24973 25654: contig of 682 bp in length
25655 25754: gap of 100 bp

25755 26510: contig of 756 bp in length
26511 26610: gap of 100 bp
26611 27369: contig of 759 bp in length
27370 27469: gap of 100 bp
27470 28209: contig of 740 bp in length
28210 28309: gap of 100 bp
28310 29077: contig of 768 bp in length
29078 29177: gap of 100 bp
29178 29924: contig of 747 bp in length
29925 30024: gap of 100 bp
30025 30796: contig of 772 bp in length
30797 30896: gap of 100 bp
30897 31655: contig of 759 bp in length
31656 31755: gap of 100 bp
31756 32502: contig of 747 bp in length
32503 32602: gap of 100 bp
32603 33369: contig of 767 bp in length
33370 33469: gap of 100 bp
33470 34233: contig of 764 bp in length
34234 34333: gap of 100 bp
34334 35081: contig of 748 bp in length
35082 35181: gap of 100 bp
35182 35951: contig of 770 bp in length
35952 36051: gap of 100 bp
36052 36820: contig of 769 bp in length
36821 36920: gap of 100 bp
36921 37694: contig of 774 bp in length
37695 37794: gap of 100 bp
37795 38563: contig of 769 bp in length
38564 38663: gap of 100 bp
38664 39427: contig of 764 bp in length
39428 39527: gap of 100 bp
39528 40285: contig of 758 bp in length
40286 40385: gap of 100 bp
40386 41148: contig of 763 bp in length
41149 41248: gap of 100 bp
41249 42007: contig of 759 bp in length
42008 42107: gap of 100 bp
42108 42867: contig of 760 bp in length
42868 42967: gap of 100 bp
42968 43716: contig of 749 bp in length
43717 43816: gap of 100 bp
43817 44549: contig of 733 bp in length
44550 44649: gap of 100 bp
44650 45404: contig of 755 bp in length
45405 45504: gap of 100 bp
45505 46270: contig of 766 bp in length
46271 46370: gap of 100 bp
46371 47132: contig of 762 bp in length
47133 47232: gap of 100 bp
47233 48006: contig of 774 bp in length
48007 48106: gap of 100 bp
48107 48875: contig of 769 bp in length
48876 48975: gap of 100 bp
48976 49735: contig of 760 bp in length
49736 49835: gap of 100 bp
49836 50585: contig of 750 bp in length
50586 50685: gap of 100 bp
50686 51449: contig of 764 bp in length
51450 51549: gap of 100 bp
51550 52306: contig of 757 bp in length
52307 52406: gap of 100 bp
52407 53158: contig of 752 bp in length
53159 53258: gap of 100 bp
53259 54002: contig of 744 bp in length
54003 54102: gap of 100 bp
54103 54883: contig of 781 bp in length
54884 54983: gap of 100 bp
54984 55748: contig of 765 bp in length
55749 55848: gap of 100 bp
55849 56621: contig of 773 bp in length
56622 56721: gap of 100 bp
56722 57477: contig of 756 bp in length


```

misc_feature      /note="assembly_fragment"
33561..42328
misc_feature      /note="assembly_fragment"
42429..50674
misc_feature      /note="assembly_fragment"
50775..59862
misc_feature      /note="assembly_fragment"
59963..71699
misc_feature      /note="assembly_fragment"
71800..82583
misc_feature      /note="assembly_fragment"
82684..95072
misc_feature      /note="assembly_fragment"
clone_end:T7
vector_side:right"
95173..108949
misc_feature      /note="assembly_fragment"
109050..126668
misc_feature      /note="assembly_fragment"
126769..156184
misc_feature      /note="assembly_fragment"
BASE COUNT      51507 a 26941 c 26653 g 49378 t 1505 others
ORIGIN

```

```

Query Match      64.0%; Score 16; DB 71; Length 156184;
Best Local Similarity 100.0%; Pred.No.12;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      7 gcaagtcacagaatla 22
|||||
Db 123532 GCAAGTCACAGAATTA 123517

```

```

RESULT 42
AL391821      158408 bp      DNA      PRI      01-NOV-2000
LOCUS
DEFINITION
AL391821      Human DNA sequence from clone RP11-212B22 on chromosome X, complete
sequence.
ACCESSION
AL391821.7      GI:11121082
VERSION
KEYWORDS
SOURCE
HMG.
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 158408)
AUTHORS
Heath,P.
JOURNAL
Direct Submission
Submitted (31-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Nov 8, 2000 this sequence version replaced gi:10944214.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C/elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at

```

```

FEATURES
source
http://www.sanger.ac.uk/HGP/ChrX
RP11-212B22 is from the library RPT-11.1 constructed at the
Roswell Park Cancer Institute by the group of Pletier de Jong. For
further details see http://bacpac.med.buffalo.edu/
VECTOR: pBAC3.6
This sequence is the entire insert of clone RP11-212B22 The true
left end of clone RP11-48719 is at 133439 in this sequence. The
true right end of clone RP11-35G18 is at 153997 in this sequence.
Location/Qualifiers
1..158408
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone_lib="RPT-11.1"
2..226
/note="LIMB2 repeat: matches 5946..6163 of consensus"
295..525
/note="MLT1H repeat: matches 229..494 of consensus"
648..677
/note="15 copies 2 mer ca 90% conserved"
2598..2649
/note="26 copies 2 mer ac 90% conserved"
2780..2853
/note="37 copies 2 mer ct 74% conserved"
complement(3446..3841)
/note="match: GSS: Em:A0136371"
4491..4655
/note="LIME repeat: matches 468..628 of consensus"
4804..5008
/note="MER58C repeat: matches 3..88 of consensus"
5044..5181
/note="LIM4 repeat: matches 1500..1639 of consensus"
5271..5414
/note="LIMEC repeat: matches 2242..2385 of consensus"
5420..5718
/note="ALUSq repeat: matches 1..298 of consensus"
5726..5994
/note="ALUo repeat: matches 1..271 of consensus"
6011..6787
/note="LIMEC repeat: matches 2356..2818 of consensus"
6801..7262
/note="LIPAS repeat: matches 5682..6143 of consensus"
7305..7750
/note="LIMAS9 repeat: matches 5660..6308 of consensus"
7799..8106
/note="LIM4 repeat: matches 2841..3351 of consensus"
8246..8761
/note="LIM4 repeat: matches 869..1405 of consensus"
8819..9030
/note="LIM4 repeat: matches 1391..1596 of consensus"
9073..10246
/note="LIPAS3 repeat: matches 4979..6146 of consensus"
10260..10753
/note="LIM4 repeat: matches 1575..1608 of consensus"
11391..11509
/note="FLAMC repeat: matches 1..132 of consensus"
11510..11821
/note="ALUSq repeat: matches 1..312 of consensus"
11834..12407
/note="LIM4 repeat: matches 2370..2908 of consensus"
12408..12684
/note="ALUo repeat: matches 1..277 of consensus"
12685..13112
/note="LIM4 repeat: matches 2908..3346 of consensus"
13191..13493
/note="LIM4 repeat: matches 3350..3653 of consensus"
13515..13538
/note="12 copies 2 mer at 100% conserved"
13539..13894
/note="LIMAS9 repeat: matches 4974..5334 of consensus"
13895..14205
/note="ALUSq repeat: matches 1..310 of consensus"

```

```

repeat_region 14206..15193
/note="L1MAY repeat: matches 5334..6305 of consensus"
repeat_region 15269..15357
/note="L2 repeat: matches 2644..2748 of consensus"
repeat_region 15388..15841
/note="L1P repeat: matches 4685..5138 of consensus"
misc_feature complement(15793..16200)
/note="match: GSS: Em:A034620"
repeat_region 16099..16221
/note="L2 repeat: matches 2572..2710 of consensus"
repeat_region 16280..16732
/note="L2 repeat: matches 1992..2403 of consensus"
repeat_region 17982..18300
/note="AlusX repeat: matches 1..307 of consensus"
repeat_region 18532..18756
/note="MIR repeat: matches 14..262 of consensus"
repeat_region 20534..20587
/note="27 copies 2 mer tt 81% conserved"
repeat_region 21039..21171
/note="AlusG/x repeat: matches 79..209 of consensus"
repeat_region 22602..22902
/note="AlusX repeat: matches 2..302 of consensus"
repeat_region 23484..23533
/note="25 copies 2 mer gt 96% conserved"
repeat_region 23532..23561
/note="10 copies 3 mer gtg 90% conserved"
repeat_region 23877..25040
/note="L2 repeat: matches 1141..2323 of consensus"
repeat_region 25257..25415
/note="L2 repeat: matches 2582..2748 of consensus"
repeat_region 26553..26594
/note="21 copies 2 mer aa 76% conserved"
misc_feature 26798..27063
/note="match: GSS: Em:A0343061"
repeat_region 27948..28106
/note="MERSB repeat: matches 10..175 of consensus"
repeat_region 28430..28618
/note="MERS3 repeat: matches 1..189 of consensus"
repeat_region 28941..29026
/note="MIR repeat: matches 112..198 of consensus"
repeat_region 30445..30809
/note="THE1B repeat: matches 1..364 of consensus"
repeat_region 31228..31341
/note="L2 repeat: matches 2629..2749 of consensus"
repeat_region 31966..32079
/note="L2 repeat: matches 2620..2698 of consensus"
repeat_region 32105..32398
/note="AluY repeat: matches 1..294 of consensus"
misc_feature complement(32313..32867)
/note="match: GSS: Em:A0357777"
misc_feature complement(32386..32876)
/note="match: GSS: Em:A0357649"
repeat_region 32611..32654
/note="22 copies 2 mer tt 79% conserved"
repeat_region 32804..32877
/note="MERSA repeat: matches 21..114 of consensus"
repeat_region 33673..33723
/note="MIR repeat: matches 102..152 of consensus"
misc_feature complement(34512..34778)
/note="match: STS: Em:LL24830"
repeat_region 35370..35678
/note="AluY repeat: matches 1..309 of consensus"
misc_feature complement(36574..36908)
/note="match: STS: Em:HSC93A6"
repeat_region 36808..37545
/note="L2 repeat: matches 2017..2749 of consensus"
repeat_region 37906..38033
/note="MERSA repeat: matches 59..189 of consensus"
repeat_region 38390..38732
/note="MIR1A1 repeat: matches 1..365 of consensus"
repeat_region 38744..38995
/note="MIR repeat: matches 8..261 of consensus"
misc_feature complement(39579..40134)

```

```

Query Match 64.0%; Score 16; DB 90; Length 158408;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 aagtcacagattatt 24
Db 117994 AAGTCACAGATTATT 118009

RESULT 43
AC022621
LOCUS
DEFINITION Homo sapiens clone RP11-23M2, WORKING DRAFT SEQUENCE, 27 unordered
pieces.
ACCESSION AC022621
VERSION AC022621.4
KEYWORDS GI:7239605
SOURCE HTG; HTGS_PHASE1; HTGS_DRAFT.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 167110)
REFERENCE 1 (bases 1 to 167110)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckwith,R., Beda,P.,
Boguslavsky,I., Bouckhalter,B., Brown,A., Burkett,G., Castle,A.,
Choquel,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeBartolano,K., Dewar,K., Domino,M., Doyle,M., Fennel,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehotzky,J., Levine,R., Liu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGuck,A., McKernan,K.,
McPheters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Tromann,N.,
Stojanovic,N., Subramanian,A., Talmes,J., Tesfaye,S., Theodore,J.,
Tirelli,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A., and Zody,M.
Direct Submission
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 14, 2000 this sequence version replaced gi:698082.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WITR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

```

```

----- Project Information
Center project name: 14396
Center clone name: 23.M2
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 154068 bases at least Q40
Consensus quality: 160659 bases at least Q30
Consensus quality: 162957 bases at least Q20
Insert size: 152000; agarose-1p
Insert size: 164510; sum-of-ctrls
Quality coverage: 4.2 in Q20 bases; agarose-1p
Quality coverage: 3.9 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1025 1124: gap of 1024 bp in length
1125 2237: contig of 1113 bp in length
2238 2337: gap of 100 bp
2338 3378: contig of 1041 bp in length
3379 3478: gap of 100 bp
3479 5102: contig of 1624 bp in length
5103 5202: gap of 100 bp
5203 7984: contig of 2782 bp in length
7985 8084: gap of 100 bp
8085 10770: contig of 2686 bp in length
10771 10870: gap of 100 bp
10871 13246: contig of 2376 bp in length
13247 13346: gap of 100 bp
13347 15161: contig of 1815 bp in length
15162 15261: gap of 100 bp
15262 18865: contig of 3604 bp in length
18866 18965: gap of 100 bp
18966 22975: contig of 4010 bp in length
22976 23075: gap of 100 bp
23076 26777: contig of 3702 bp in length
26778 26877: gap of 100 bp
26878 31543: contig of 4666 bp in length
31544 31643: gap of 100 bp
31644 36119: contig of 4476 bp in length
36120 36219: gap of 100 bp
36220 40123: contig of 3904 bp in length
40124 40223: gap of 100 bp
40224 42742: contig of 2519 bp in length
42743 42842: gap of 100 bp
42843 49609: contig of 6767 bp in length
49610 49709: gap of 100 bp
49710 55228: contig of 5519 bp in length
55229 55328: gap of 100 bp
55329 60561: contig of 5233 bp in length
60562 60661: gap of 100 bp
60662 67385: contig of 6724 bp in length
67386 67485: gap of 100 bp
67486 73407: contig of 5922 bp in length
73408 73507: gap of 100 bp
73508 81201: contig of 7694 bp in length
81202 81301: gap of 100 bp
81302 89851: contig of 8550 bp in length
89852 89951: gap of 100 bp
89952 97520: contig of 7569 bp in length
97521 97620: gap of 100 bp
97621 110089: contig of 12469 bp in length
110090 110189: gap of 100 bp
110190 122757: contig of 12568 bp in length
122758 122857: gap of 100 bp

```

```

FEATURES
source
* 122858 143904: contig of 21047 bp in length
* 143905 144004: gap of 100 bp
* 144005 167110: contig of 23106 bp in length.
Location/Qualifiers
1. 167110
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-23M2"
/clone_lib="RPC1-11 Human Male BAC"
1. 1024
/note="assembly-fragment"
1125..2237
/note="assembly-fragment"
2338..3378
/note="assembly-fragment"
3479..5102
/note="assembly-fragment"
5203..7984
/note="assembly-fragment"
8085..10770
/note="assembly-fragment"
10871..13246
/note="assembly-fragment"
13347..15161
/note="assembly-fragment"
vector_end:r7
vector_side:right"
15262..18865
/note="assembly-fragment"
18966..22975
/note="assembly-fragment"
23076..26777
/note="assembly-fragment"
26878..31543
/note="assembly-fragment"
31644..36119
/note="assembly-fragment"
36220..40123
/note="assembly-fragment"
40224..42742
/note="assembly-fragment"
42843..49609
/note="assembly-fragment"
49710..55228
/note="assembly-fragment"
55329..60561
/note="assembly-fragment"
60662..67385
/note="assembly-fragment"
67486..73407
/note="assembly-fragment"
73508..81201
/note="assembly-fragment"
81302..89851
/note="assembly-fragment"
89952..97520
/note="assembly-fragment"
vector_end:sp6
vector_side:left"
97621..110089
/note="assembly-fragment"
110190..122757
/note="assembly-fragment"
122858..143904
/note="assembly-fragment"
144005..167110
/note="assembly-fragment"
BASE COUNT 48189 a 34563 c 33851 g 47895 t 2612 others
ORIGIN
Query Match 64.0%; Score 16; DB 67; Length 167110;
Best local Similarity 100.0%; Pred. No. 12;

```

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 agtccagaaattattt 25
 |||||||||||||||
 Db 49204 AGTCACAGAAATTATT 49219

RESULT 44
 AC024560 167176 bp DNA HTG 16-APR-2001
 LOCUS Homo sapiens chromosome 3 clone RP11-496H1, WORKING DRAFT SEQUENCE,
 DEFINITION 8 unordered pieces.
 AC024560
 AC024560.15 GI:13624340
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FOLDTOP.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 167176)
 Wuziy,D.M., Adams,C., Adio-Oduola,B., Ali-oshman,F.R., Allen,C.,
 Alstbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,
 Benton,J., Bimange,K., Blankenburg,K., Bonini,D., Bouck,J.,
 Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carroon,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Dem,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunatane,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
 Hollins,B., Homsi,F., Howard,S., Huber,D., Hulyk,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korah,J.,
 Kovar,C., Kratochvic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
 Louisedge,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
 Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
 Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Monabbat,K.,
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newman,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
 Ogutu,M., Okunolu,G., Otagunye,N., Oviedo,R., Pace,A., Payton,B.,
 Peetly,O., Perez,L., Peters,L., Plickens,R., Prius,E., Pu,L.L.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
 Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshitari,N.,
 Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
 Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
 Watlington,S., Williams,G., Williamson,A., Wlezyk,R., Woodson,S.,
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 167176)
 Worley,K.C.
 Direct Submission
 Submitted (29-FEB-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Apr 14, 2001 this sequence version replaced gi:13374578.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information
 Center project name: HAW
 Center clone name: RP11-496H1
 ----- Summary Statistics
 Sequencing vector: Plasmid; M77789
 Sequencing vector: M13; L08821
 Chemistry: Dye-Primer Body: 6% of reads
 Chemistry: Dye-terminator Big Dye: 94% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 167785 bases at least Q40
 Consensus quality: 165676 bases at least Q30
 Consensus quality: 167353 bases at least Q20
 Estimated insert size: 164567; sum-of-coverage estimation
 Quality coverage: 0x in Q20 bases; agarose-IP estimation
 Quality coverage: 8.9x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 69736: contig of 69736 bp in length
 * 69737 69836: gap of unknown length
 * 69837 101725: contig of 31889 bp in length
 * 101726 101825: gap of unknown length
 * 101826 122611: contig of 20786 bp in length
 * 122612 122711: gap of unknown length
 * 122712 139512: contig of 16701 bp in length
 * 139513 139512: gap of unknown length
 * 139513 154634: contig of 15122 bp in length
 * 154635 154734: gap of unknown length
 * 154735 164776: contig of 10042 bp in length
 * 164777 164876: gap of unknown length
 * 164877 166032: contig of 1156 bp in length
 * 166033 166132: gap of unknown length
 * 166133 167176: contig of 1044 bp in length.
 * Location/Qualifiers
 1. 167176
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clone="RP11-496H1"

BASE COUNT 45856 a 40392 c 37326 g 42895 t 707 others
 ORIGIN

Query Match 64.0%; Score 16; DB 68; Length 167176;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 agtccagaaattattt 25
 |||||||||||||||
 Db 167115 AGTCACAGAAATTATT 167130

RESULT 45
 AC010816 167803 bp DNA HTG 10-SEP-2000
 LOCUS Homo sapiens clone RP11-3C13, WORKING DRAFT SEQUENCE, 11 unordered
 DEFINITION pieces.
 AC010816
 AC010816.4 GI:10047737
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 167803)

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome, clone RP11-3G13
Unpublished
2 (bases 1 to 167803)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckert,Y., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,C., Hagos,B., Heaford,A., Horton,L.,
Howard,J.C., Johnson,R., Jones,C., Kam,L., Karacas,A., Klein,J.,
Lehoczky,J., Lien,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,A., Talamas,J.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT
Submitted (23-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 10, 2000 this sequence version replaced g1:7637262.
All repeats were identified using RepeatMasker:
Smtl, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIRK

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: 12730

Center clone name: 3.G.13

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 150949 bases at least Q40

Consensus quality: 159528 bases at least Q30

Consensus quality: 163392 bases at least Q20

Insert size: 14500; agarose-rp

Insert size: 166803; sum-of-contigs

Quality coverage: 4.0 in Q20 bases; agarose-rp

Quality coverage: 3.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1593: contig of 1593 bp in length
* 1594 1693: gap of 100 bp
* 1694 2814: contig of 1121 bp in length
* 2815 2914: gap of 100 bp
* 2915 47564: contig of 44650 bp in length
* 47565 47664: gap of 100 bp
* 47665 59530: contig of 11866 bp in length
* 59531 59630: gap of 100 bp
* 59631 70197: contig of 10567 bp in length
* 70198 70297: gap of 100 bp
* 70298 81312: contig of 11015 bp in length
* 81313 81412: gap of 100 bp
* 81413 99699: contig of 18287 bp in length
* 99700 99799: gap of 100 bp
* 99800 116730: contig of 16931 bp in length
* 116731 116830: gap of 100 bp
* 116831 139677: contig of 22847 bp in length
* 139678 139777: gap of 100 bp
* 139778 165369: contig of 25592 bp in length
* 165370 165469: gap of 100 bp

FEATURES * 165470 167803: contig of 2334 bp in length.
location/Qualifiers
source

1..167803
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-3G13"
/clone_1fb="RPC1-11 Human Male BAC"
1..1593

misc_feature
/note="assembly_fragment"
clone_end:SP6
vector_side:left

misc_feature
1694..2814
/note="assembly_fragment"
2915..47564
/note="assembly_fragment"

misc_feature
47665..59530
/note="assembly_fragment"
59631..70197
/note="assembly_fragment"

misc_feature
70298..81312
/note="assembly_fragment"
81413..99699
/note="assembly_fragment"

misc_feature
99800..116730
/note="assembly_fragment"
116831..139677
/note="assembly_fragment"

misc_feature
139778..165369
/note="assembly_fragment"
165470..167803
/note="assembly_fragment"
clone_end:T7
vector_side:right

BASE COUNT 54453 a 30838 c 30293 g 51217 t 1002 others
ORIGIN
Query Match 64.0%; Score 16; DB 61; Length 167803;
Best Local Similarity 100.0%; Pred. NO. 12;
Matches 16; Conservative 0; Mismatches 0; Gaps 0;

QY 10 agtcacagaattattt 25
|||||
DB 83723 AGTCACAGATTATT 83738

Search completed: October 9, 2001, 15:44:37
Job time: 14663 sec

Wed Oct 10 07:45:39 2001

us-09-396-196f-2.oli.rge

Page 46

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:51:43 ; Search time 469.84 Seconds
(without alignments)
33.410 Million cell updates/sec

Title: US-09-396-196f-2

Perfect score: 25
Sequence: 1 attgcgcagtcacagattattt 25

Scoring table: OLIGO_MUC
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 9

Total number of hits satisfying chosen parameters: 27371

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

N.GeneSeq_0601.*
1: /SIDSL/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseqn/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	20	AA01303
2	25	100.0	1084	10	AA01329
3	25	100.0	1121	7	AA060496
4	25	100.0	5872	15	AA062386
5	17	68.0	561	17	AA056307
6	15	60.0	561	22	AA094235
7	15	60.0	1771	20	AA030602
8	15	60.0	3520	22	AA093909
9	15	60.0	3520	22	AA029361
10	15	60.0	116624	19	AA052850
11	14	56.0	329	21	AA011925

12	14	56.0	360	22	AA065421
13	14	56.0	385	14	AA039649
14	14	56.0	385	14	AA059061
15	14	56.0	902	21	AA056756
16	14	56.0	1050	9	AA080927
17	14	56.0	1677	17	AA031932
18	14	56.0	1787	17	AA031928
19	14	56.0	1890	20	AA052940
20	14	56.0	2233	19	AA010120
21	14	56.0	2585	22	AA077541
22	14	56.0	2845	21	AA069462
23	14	56.0	3413	20	AA072330
24	14	56.0	5684	21	AA059953
25	14	56.0	5773	17	AA076014
26	14	56.0	8041	17	AA010105
27	14	56.0	32351	21	AA021307
28	14	56.0	32351	21	AA035185
29	14	56.0	40298	21	AA021311
30	14	56.0	40298	21	AA035189
31	14	56.0	81001	22	AA030035
32	14	56.0	1038602	20	AA0201425
33	13	52.0	87	16	AA022576
34	13	52.0	105	16	AA020380
35	13	52.0	167	16	AA020233
36	13	52.0	187	21	AA026213
37	13	52.0	354	15	AA077082
38	13	52.0	363	21	AA050159
39	13	52.0	378	13	AA025667
40	13	52.0	387	21	AA079285
41	13	52.0	408	21	AA074604
42	13	52.0	409	21	AA03836
43	13	52.0	415	21	AA056783
44	13	52.0	441	21	AA07649
45	13	52.0	476	21	AA07951

ALIGNMENTS

RESULT 1	
AA01303	AA01303 standard; DNA; 1041 BP.
XX	XX
AC	AA01303;
XX	XX
DT	12-APR-1999 (first entry)
XX	XX
DE	E. coli biotin synthetase (Biot) coding sequence.
XX	XX
KM	DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW	biotin synthetase; biotin production; vitamin H; Biot; ss.
XX	XX
OS	Escherichia coli.
XX	XX
PN	US0869719-A.
XX	XX
PD	09-FEB-1999.
XX	XX
PF	30-APR-1997; 97US-0846338.
XX	XX
PR	30-APR-1997; 97US-0846338.
PR	08-MAR-1995; 95US-0401068.
XX	XX
PA	(NOVS) NOVARTIS FINANCE CORP.
PI	Patton DA;
XX	XX
DR	WPI: 1999-152902/13.
XX	XX
PT	P-PSDB; AA073906.
PT	Transgenic plants with high biotin levels - transformed with DNA
PT	encoding di:amino-pelargonic acid amino-transferase or biotin
PT	synthase

Novel human polynu
Expressed Sequence
Human brain Expres
Human transmembran
Lividans Exported
Retinoid X recepto
Retinoid X recepto
Human prostate tum
Human retinoid rec
Caenorhabditis bri
Human secreted pro
F. balustium CP70
Human secreted pro
Human ORX ORF1569
Adherence conferti
Human low adenosin
Human adenosine re
Human low adenosin
Human adenosine re
Human apolipoprote
Complete genome se
Human gene signatu
Human gene signatu
Human gene signatu
Human secreted pro
Human genome fragm
Antibody 5H7 heavy
Sequence of the an
Human lung tumour-
Murine 5B3 antibod
Human secreted pro
Eucalyptus grandis
Human secreted pro
Human secreted pro

KW dethiobiotin synthase; DAPA synthase;
 KM S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
 KM seborrhea; dermatitis; ds.
 OS Escherichia coli DSM498.
 XX
 Key Location/Qualifiers
 FH 1..96
 FT promoter /tag= a
 FT /function= "promoter plac"
 FT /evidence= EXPERIMENTAL
 FT 23..28
 FT -35_signal /tag= b
 FT /standard_name= "promoter plac"
 FT 45..50
 FT -10_signal /tag= c
 FT /evidence= EXPERIMENTAL
 FT /standard_name= "promoter plac"
 FT 105..109
 FT RBS /tag= d
 FT /evidence= EXPERIMENTAL
 FT /standard_name= "bioB RBS no. 9"
 FT 117..1157
 FT CDS /tag= e
 FT /product= "biotin synthase"
 FT /evidence= EXPERIMENTAL
 FT /gene= "bioB"
 FT /number= 1
 FT 1141..1146
 FT RBS /tag= f
 FT /standard_name= "bioF RBS"
 FT 1154..2311
 FT CDS /tag= g
 FT /RC_number= 2.3.1.47
 FT /product= "KAPA synthase"
 FT /evidence= EXPERIMENTAL
 FT /gene= "bioF"
 FT /number= 2
 FT /standard_name= "8-amino-7-oxononanoate synthase"
 FT 2284..2288
 FT RBS /tag= h
 FT /standard_name= "bioC RBS"
 FT 2295..3050
 FT CDS /tag= i
 FT /function= "involved in pimeloyl-CoA synthesis"
 FT /product= "protein"
 FT /gene= "bioC"
 FT /number= 3
 FT 3030..3033
 FT RBS /tag= j
 FT /standard_name= "bioD RBS"
 FT 3043..3753
 FT CDS /tag= k
 FT /RC_number= 6.3.3.3
 FT /product= "DTB synthase"
 FT /evidence= EXPERIMENTAL
 FT /gene= "bioD15"
 FT /number= 4
 FT /standard_name= "dethiobiotin synthase"
 FT 3712..3750
 FT misc_RNA /tag= l
 FT /note= "bioD15 substitution"
 FT 3742..3746
 FT RBS /tag= m
 FT /standard_name= "bioA RBS"
 FT 3750..5039
 FT CDS /tag= n
 FT /RC_number= 2.6.1.62
 FT /product= "DAPA synthase"
 FT /evidence= EXPERIMENTAL
 FT /gene= "bioA"
 FT /number= 5
 FT /standard_name= "S-adenosyl-L-methionine: 8-amino-

FT RBS 5088..5093 7-oxononanoate aminotransferase"
 FT /tag= o
 FT /standard_name= "ORF1 RBS"
 FT 5098..5574
 FT CDS /tag= p
 FT /function= "unknown, involved in biotin synthesis"
 FT /product= "protein"
 FT /evidence= EXPERIMENTAL
 FT /gene= "ORF1"
 FT /number= 6
 FT 5583..5644
 FT terminator /tag= q
 FT /standard_name= "rho-independent transcriptional
 FT stem_loop terminator"
 FT 5583..5605
 FT /tag= r
 FT
 PD WO9408023-A.
 XX 14-APR-1994.
 XX
 PF 01-OCT-1993; 93WO-EP02688.
 XX
 PR 02-OCT-1992; 92CH-0003124.
 PR 15-JUL-1993; 93CH-0002134.
 XX
 PA (LONZ) LONZA AG.
 XX
 PI Birch O, Brass J, Fuhrmann M, Shaw N;
 XX
 DR WPI: 1994-135587/16.
 DR P-PDB: AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
 XX
 PT Biotechnological biotin prodn. using enterobacterial biotin-gene
 PT - providing vitamin H in high yield
 XX
 PS Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
 XX
 CC The sequence is derived from plasmid pB030A-15/9 contg. the
 CC bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
 CC of biotin, arranged in a transcription unit. Microorganisms
 CC confg. these DNA fragments or plasmids may be used in the prodn.
 CC of biotin. Biotin (Vitamin H) may prevent seborrhea, dermatitis,
 CC loss of appetite and tiredness.
 CC
 SQ Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;
 XX
 Query Match 100.0%; Score 25; DB 15; Length 5872;
 Best local Similarity 100.0%; Pred. NO. 1..3e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 atgtgcgaagtcacagaattatt 25
 ||||||||||||||||||||
 Db 140 atgtgcgaagtcacagaattatt 164
 RESULT 5
 AA056307/c
 ID AA056307 standard; DNA; 17 BP.
 XX
 AC AA056307;
 XX
 DT 13-APR-1994 (first entry)
 XX
 DE BioB DNA primer.
 XX
 KW BioA; BioB; promoter; biotin; operon; primer; ss.
 XX
 OS Synthetic.
 XX
 PN JP05219956-A.

XX 31-AUG-1993.
 XX 14-SEP-1992: 92JP-0244792.
 XX 14-SEP-1992: 92JP-0244792.
 XX (SHIS) SHISEIDO CO LTD.
 XX (TAKE) TAKEDA CHEM IND LTD.
 XX WPI: 1993-308323/39.
 XX DNA sequence of biotin operon - has base sequence of E. coli
 PT mutated by base pair(s) compared to wild type
 XX Example 1; Fig 8; 11pp; Japanese.
 CC A novel DNA sequence comprises the E.coli biotin operon (BO) in which
 CC the control region of BO or the region near the biob initiation
 CC codon is mutated by at least one base pair compared to its
 CC wild type. Two primers (AA056306-Q56307) are described in Example 1.
 CC A microorganism belonging to Escherichia genus, transformed by
 CC a recombinant plasmid carrying such DNA can be used for the prodn.
 CC of biotin-active substances.
 XX Sequence 17 BP; 3 A; 4 C; 4 G; 6 T; 0 other;

Query Match 68.0%; Score 17; DB 14; Length 17;
 Best Local Similarity 100.0%; Pred. No. 0.51;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gtcgaagtcacagaat 20
 |||||
 DB 17 GTCGAAGTCACAGAAT 1

RESULT 6
 AAF94235
 ID AAF94235 standard; DNA; 561 BP.

AC AAF94235;
 XX 23-MAY-2001 (first entry)

DE Primer specific for DNA encoding secretory/membrane protein SEQ ID 669.

KW Human; secretory protein; membrane protein; vaccine; gene therapy;
 KW rheumatoid arthritis; diabetes; PCR primer; ss.

OS Synthetic.

PN EPI067182-A2.

PD 10-JAN-2001.

PE 07-JUL-2000; 2000EP-0114090.

PR 08-JUL-1999; 99JP-0194179.

PR 11-JAN-2000; 2000JP-0118775.

PR 02-MAY-2000; 2000JP-0183766.

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
 DR WPI; 2001-093989/11.

PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
 PT gene therapy or as candidate target molecules in drug development -
 XX Claim 5; SEQ ID 669; 609pp+ CD ROM; English.

CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
 CC which encode human secretory or membrane proteins represented by
 CC AAF88317 - AAF88419. Included in the invention are primers
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
 CC cDNA sequences of the invention. The invention also includes methods for
 CC the production of antibodies directed against the proteins, and cDNA
 CC sequences, which can be used in vaccines. The polynucleotide sequences
 CC can be used in gene therapy. The polynucleotide sequences and the
 CC proteins they encode may be used in the prevention, treatment and
 CC diagnosis of diseases associated with inappropriate secretory
 CC protein/membrane protein expression. The nucleic acids and complementary
 CC sequences may also be used as DNA probes in diagnostic assays
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
 CC presence of similar nucleic acid sequences in samples. They may also be
 CC used to study the expression and function of secretory proteins/membrane
 CC polypeptides and their role in metabolism. The polypeptides may be used
 CC as antigens in the production of antibodies against them and in assays to
 CC identify modulators (agonists and antagonists) of expression and
 CC activity. The antibodies and antagonists may also be used as therapeutic
 CC agents to down regulate expression and activity. The antibodies may also
 CC be used as diagnostic agents for detecting the presence of the
 CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay
 CC (ELISA)). Examples of diseases which may be treated include rheumatoid
 CC arthritis and diabetes.
 XX Sequence 561 BP; 213 A; 83 C; 114 G; 149 T; 2 other;

Query Match 60.0%; Score 15; DB 22; Length 561;
 Best Local Similarity 100.0%; Pred. No. 6.9;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 caagtcacagaatta 22
 |||||
 DB 339 caagtcacagaatta 353

RESULT 7
 AA230602
 ID AA230602 standard; DNA; 1771 BP.

AC AA230602;

DE 18-JAN-2000 (first entry)

DE Mouse integrin alpha 4 coding sequence.

KW Human; integrin; antisense; oligonucleotide; inhibition; expression;
 KW very late antigen; CD49d; CD29; cell surface; leucocyte; adhesion;
 KW vascular endothelial cell; vascular endothelium; migration; inflammation;
 KW atherosclerosis; allergy; asthma; rheumatoid arthritis; tumor;
 KW metastasis; circulatory system; autoimmune disease; Grave's disease;
 KW Hashimoto's thyroiditis; encephalomyelitis; multiple sclerosis; ds.

OS Mus sp.

PN US5968826-A.

PD 19-OCT-1999.

PE 05-OCT-1998; 98US-0166203.

PR 05-OCT-1998; 98US-0166203.

PR 05-OCT-1998; 98US-0166203.

PA (ISIS-) ISIS PHARM INC.

PI Bennett CF, Cowser LM, Condon TP;
 DR WPI: 1999-590416/50.

DR P-PSDB; AAY44154.

PT Antisense inhibition of integrin alpha4 expression useful for treating
 PT inflammatory diseases such as atherosclerosis, allergies, asthma and

PT arthritis -
 XX
 PS Example 12: Column 71-74: 40pp: English.
 CC This sequence represents the coding region of the mouse Integrin
 CC alpha4 gene. The invention relates to the generation of antisense
 CC oligonucleotides targeted to the integrin alpha4 gene which are used
 CC for inhibiting expression of the integrin alpha4 mRNA or protein.
 CC Integrin alpha4 is a component of Very Late Antigen (VLA)-4 (also
 CC called alpha4beta1 and CD49/CD29). VLA-4 is expressed on the cell
 CC surfaces of leucocytes and vascular endothelial cells and mediates the
 CC adhesion of leucocytes to the vascular endothelium prior to migration
 CC into the surrounding tissues. This migration is an essential step in
 CC inflammation and hence VLA-4 (and consequently integrin alpha4) is a
 CC potential therapeutic target for treating inflammatory diseases and
 CC the damaging effects of excessive inflammation. These disorders include
 CC atherosclerosis, allergies, asthma, rheumatoid arthritis and tumor cell
 CC metastasis (VLA-4 is involved in migration of the tumor cells through
 CC the extracellular matrix into the circulatory system). VLA-4 is also
 CC involved in a number of autoimmune diseases such as Grave's disease,
 CC Hashimoto's thyroiditis, encephalomyelitis (EAE), multiple sclerosis,
 CC VLA-4 may also be involved in promoting adhesion (i.e. retention) of
 CC hematopoietic stem cells in bone-marrow and in allograft rejection.
 CC
 XX
 SQ Sequence 1771 BP; 371 A; 471 C; 519 G; 410 T; 0 other;

Query Match 60.0%; Score 15; DB 20; Length 1771;
 Best Local Similarity 100.0%; Pred. No. 6.9;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 7 gcaagtcacagaatt 21
 |||||
 Db 312 gcaagtcacagaatt 326

RESULT 8
 AAF93909/c
 ID AAF93909 standard; cDNA; 3520 BP.
 XX
 AC AAF93909;
 XX
 DT 23-MAY-2001 (first entry)
 XX
 DE Human cDNA encoding a membrane or secretory protein clone PSRC0256.
 XX
 KW Human; secretory protein; membrane protein; vaccine; gene therapy;
 KW rheumatoid arthritis; diabetes; ss.
 XX
 OS Homo sapiens.
 XX
 PN EPI067182-A2.
 XX
 PD 10-JAN-2001.
 XX
 PF 07-JUL-2000; 2000EP-0114090.
 XX
 PR 08-JUL-1999; 99JP-0194179.
 PR 11-JAN-2000; 2000JP-0118775.
 PR 02-MAY-2000; 2000JP-0183766.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
 DR WPI: 2001-093989/11.
 DR P-PSDB: AAB88482.
 XX
 PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
 PT gene therapy or as candidate target molecules in drug development -
 PS Claim 1; SEQ ID 331; 609pp + CD ROM; English.
 XX

CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
 CC which encode human secretory or membrane proteins represented by
 CC AAB88317 - AAB88419. Included in the invention are primers
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
 CC cDNA sequences of the invention. The invention also includes methods for
 CC the production of antibodies directed against the proteins, and cDNA
 CC sequences, which can be used in vaccines. The polynucleotide sequences
 CC can be used in gene therapy. The polynucleotide sequences and the
 CC proteins they encode may be used in the prevention, treatment and
 CC diagnosis of diseases associated with inappropriate secretory
 CC protein/membrane protein expression. The nucleic acids and complementary
 CC sequences may also be used as DNA probes in diagnostic assays
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
 CC presence of similar nucleic acid sequences in samples. They may also be
 CC used to study the expression and function of secretory proteins/membrane
 CC polypeptides and their role in metabolism. The polypeptides may be used
 CC as antigens in the production of antibodies against them and in assays to
 CC identify modulators (agonists and antagonists) of expression and
 CC activity. The antibodies and antagonists may also be used as therapeutic
 CC agents to down regulate expression and activity. The antibodies may also
 CC be used as diagnostic agents for detecting the presence of the
 CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay
 CC (ELISA). Examples of diseases which may be treated include rheumatoid
 CC arthritis and diabetes.
 CC
 XX
 SQ Sequence 3520 BP; 871 A; 896 C; 901 G; 852 T; 0 other;

Query Match 60.0%; Score 15; DB 22; Length 3520;
 Best Local Similarity 100.0%; Pred. No. 6.9;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 8 caagtcacagaatt 22
 |||||
 Db 3182 CAAGTCACAGATT 3168

RESULT 9
 AAF29361/c
 ID AAF29361 standard; DNA; 3520 BP.
 XX
 AC AAF29361;
 XX
 DT 20-APR-2001 (first entry)
 XX
 DE Amyloid-beta protein agglutination regulating factor DNA SEQ ID 9.
 XX
 KW Human; amyloid-beta protein; agglutination regulatory factor;
 KW Alzheimer's disease; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200104299-A1.
 XX
 PD 18-JAN-2001.
 XX
 PF 06-JUL-2000; 2000WO-JP04515.
 XX
 PR 08-JUL-1999; 99JP-0194179.
 PR 18-OCT-1999; 99US-0159586.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Kawai Y, Yamazaki M, Satoh S;
 PI Arakawa H, Morita M;
 DR WPI: 2001-138347/14.
 DR P-PSDB: AAB49771.
 XX
 PT Polynucleotide encoding Amyloid-beta protein agglutination-controlling
 PT factor, useful for inhibiting or promoting agglutination or
 PT sedimentation of amyloid-beta protein and in diagnosis and screening
 PT drugs for Alzheimer's disease -

XX PS Claim 1; Page 58-63; 72pp; Japanese.
 XX CC This invention relates to polynucleotides AAF29357 - AAF29361 which
 CC encode proteins AAB49767 - AAB49771. The proteins inhibit or promote the
 CC agglutination of amyloid beta protein. The protein and polynucleotide
 CC sequences are useful in the diagnosis of Alzheimer's disease. They are
 CC also useful for screening drugs which are useful for treating Alzheimer's
 CC disease.
 XX SQ Sequence 3520 BP; 871 A; 896 C; 901 G; 852 T; 0 other;

Query Match 60.0%; Score 15; DB 22; Length 3520;
 Best Local Similarity 100.0%; Pred. No. 6.9;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 caagtcacagaattat 22
 |||||
 Db 3182 CAAGTCACAGAATTAT 3168

RESULT 10
 AAV52850/C
 ID AAV52850 standard; DNA; 116624 BP.
 XX AC AAV52850;
 XX DT 06-NOV-1998 (first entry)
 XX DE Human eyal gene contig 4405-9480.
 XX KW Human; eyal; branchio-oto-renal syndrome; BOR; carcinogenesis; cancer;
 KW kidney tumorigenesis; tumour; diagnosis; ss.
 XX OS Homo sapiens.
 XX PN WO9832849-A2.
 XX PD 30-JUL-1998.
 XX PF 28-JAN-1998; 98WO-EP00433.
 XX PR 29-JAN-1997; 97US-0036579.
 XX PA (INSP) INST PASTEUR.
 XX PI Abdelhak S, Compain S, Petit C, Vasiliki K, Vincent C;
 PI Weil D;
 XX DR WPI; 1998-427945/36.
 XX PT Nucleic acid corresponding to human genes implicated in
 PT branchio-oto-renal syndrome - useful for, e.g. diagnosis and
 PT treatment of syndrome and possibly some carcinogenic processes,
 PT particularly in kidneys
 XX PS Claim 1; Page 44-105; 191pp; English.
 XX CC The present sequence represents a portion of the human eya-1 gene.
 CC Alterations in the eyal gene are associated with branchio-oto-renal
 CC (BOR) syndrome, including the form (BO) without renal anomalies, and
 CC possibly in some carcinogenic processes, particularly in the kidney.
 CC The related genes eya-2 and eya-3 may also be involved in development
 CC of tumours and cell differentiation. Antibodies specific for EYA
 CC proteins are used to detect the proteins by immunosassay, while genetic
 CC alterations linked to BOR are detected by amplifying DNA with primers
 CC (see AAV52857 to AAV52893) then identifying any mutations or deletions.
 CC EYA1 and EYA1-B proteins, corresponding nucleic acid, antisense sequences
 CC and Ab can be used therapeutically to modulate expression of EYA1(B), or
 CC their active derivatives, especially for treating renal disease
 CC associated with eyal abnormalities, also for ensuring correct
 CC development of grafted organs, especially where these are embryonic.

CC CC Therapeutic proteins are administered to maintain 0.1-10 mu g /ml in
 CC body fluids, locally or systemically. Nucleic acid is administered in
 CC usual gene therapy vectors (optionally to cells ex vivo) at doses of
 CC 0.1-100 mu g.
 XX SQ Sequence 116624 BP; 35549 A; 20868 C; 23043 G; 37152 T; 12 other;

Query Match 60.0%; Score 15; DB 19; Length 116624;
 Best Local Similarity 100.0%; Pred. No. 6.8;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 gtccacagaattattt 25
 |||||
 Db 19735 GTCCACAGATTATT 19721

RESULT 11
 AAA41925
 ID AAA41925 standard; cDNA; 329 BP.
 XX AC AAA41925;
 XX DT 21-AUG-2000 (first entry)
 XX DE Human secreted expressed sequence tag SEQ ID NO:665.
 XX KW Human; mouse; xenopus; rat; secreted expressed sequence tag; SEST;
 KW expressed sequence tag; EST; probe; chemotactic; proliferative;
 KW immunomodulatory; haematopoietic; chemokine; analgesic; haemostatic;
 KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
 KW antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;
 KW antitumor; osteopathic; neuroprotective; nootropic; antipsoriatic;
 KW cerebroprotective; anticoagulant; antidepressant; gene therapy;
 KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
 KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
 KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
 KW central nervous system disorder; Alzheimer's disease; stroke;
 KW Parkinson's disease; Huntington's disease; coagulation disorder;
 KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
 KW tumour; infection; depression; psoriasis; ss.
 XX OS Homo sapiens.
 XX PN WO200021990-A1.
 XX PD 20-APR-2000.
 XX PF 15-OCT-1999; 99WO-US24205.
 XX PR 15-OCT-1998; 98US-0104435.
 XX PA (GENM) GENETICS INST INC.
 XX PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M;
 XX DR WPI; 2000-317937/27.
 XX PT Isolated polynucleotides, and encoded proteins, comprising secreted
 PT expressed sequence tags (SESTs), useful for treating various disorders -
 PT such as autoimmune, infectious, and central nervous system disorders -
 XX PS Claim 1; Page 308; 618pp; English.
 XX CC AAA41261 to AAA43419 represent specifically claimed secreted expressed
 CC sequence tags (SESTs), isolated from human, mouse, xenopus and rat on
 CC tissue sources. The SESTs can have a range of activities depending on
 CC the tissues they were isolated from. The activities include:
 CC chemotactic; proliferative; immunomodulatory; haematopoietic;
 CC chemokine; analgesic; haemostatic; thrombolytic; antiinflammatory;
 CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
 CC antitumour; vulnary; antitumor; osteopathic; neuroprotective;

CC nontropic; antiparkinsonian; antipsoriatic; cerebroprotective;
 CC anticonvulsant; and antidepressant. The SESTs can be used for gene
 CC therapy and in vaccines. The SESTs are useful as probes for the
 CC identification and isolation of full-length cDNAs and genomic DNA
 CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
 CC are useful in assays for determining biological activity and raising
 CC antibodies. They may be useful for treatment of autoimmune disorders
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
 CC osteoporosis, osteoarthritis, central nervous system disorders
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
 CC disease), tumours, bacterial, fungal or viral infections, depression and
 CC psoriasis. AAA43420 to AAA43425 represent linker variants which are given
 CC in the exemplification of the present invention.

XX Sequence 329 BP; 97 A; 84 C; 67 G; 81 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 329;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 tcacagaattatt 25
 |||||
 Db 198 tcacagaattatt 211

RESULT 12

AAF65421
 ID AAF65421 standard; cDNA; 360 BP.

AC AAF65421;

DT 09-APR-2001 (first entry)

DE Novel human polynucleotide, SEQ ID NO: 1177.

KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
 KM breast cancer; lung cancer; cancer detection; ss.

OS Homo sapiens.

PN WO200102568-A2.

PD 11-JAN-2001.

PF 30-JUN-2000; 2000WO-US18374.

PR 02-JUL-1999; 99US-0142310.

PR 02-JUL-1999; 99US-0142311.

PA (CHIR) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
 PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Dimaac R;

PI Cirenjakov R, Dimaac S, Dickson M, Labat I, Leshkowitz D;

PI Kita D, Garcia V, Jones LM, Strache-Crain B;

DR WPI: 2001-091805/10.

XX Library of polynucleotides for diagnosing a cancerous state of a
 PT mammalian cell and detecting cancer, particularly of the colon or
 PT prostate, comprises 3351 human polynucleotide sequences -

XX Claim 9; Page 712; 1046pp; English.

CC The present sequence is one of 3351 sequences in a library of human
 CC polynucleotides. The library is used to detect differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell and can
 CC detect colon, prostate, breast and lung cancer. The library can be used
 CC to produce probes for detection of mRNA and to produce additional copies

CC of the polynucleotides. The probes can be used for chromosome mapping of
 CC the polynucleotide and for detection of transcription levels. Ribozymes
 CC or antisense oligonucleotides can be generated. The polynucleotides and
 CC their gene products are used as genetic or biochemical markers (e.g. in
 CC blood or tissues) that will detect the earliest changes along the
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and
 CC preventive interventions. The polynucleotides, polypeptides and
 CC antibodies against them can be used in pharmaceutical compositions to
 CC treat the cancers and proliferative disorders such as neoplasia,
 CC dysplasia and hyperplasia.

XX Sequence 360 BP; 95 A; 70 C; 69 G; 125 T; 1 other;

Query Match 56.0%; Score 14; DB 22; Length 360;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 gtcaacagaattatt 24
 |||||
 Db 71 gtcaacagaattatt 84

RESULT 13

AAQ39649
 ID AAQ39649 standard; DNA; 385 BP.

AC AAQ39649;

DT 20-MAY-1993 (first entry)

DE Expressed Sequence Tag human gene marker EST00025.

KW expressed sequence tag; human genome project; chromosome;
 KM human gene sequencing; PCR mapping; somatic cell hybrids;
 KW sublocalisation; gene tagging; tissue typing.

OS Synthetic.

PN W09300353-A.

PD 07-JAN-1993.

PF 19-JUN-1992; 92WO-US05222.

PR 20-JUN-1991; 91US-0716831.

PR 12-FEB-1992; 92US-0837195.

PA (USSH) US DEPT HEALTH & HUMAN SERVICE.

PI Adams MD, Venter JC;

PI WPI: 1993-036325/04.

PT Particular expressed sequence tags from human cDNA - corresponds
 PT to transcription prods. of genes, useful for tagging genes,
 PT mapping chromosomes and tissue typing

PS Claim 3; Page 86; 199pp; English.

XX This sequence represents an EST (expressed sequence tag) ESTs are
 CC markers for human genes actually transcribed in vivo. Unlike the random
 CC genomic DNA sequence tagged sites (STSs), ESTs point directly to
 CC expressed genes. The use of ESTs could facilitate the tagging of most
 CC expressed human genes within a few years at a fraction of the cost of
 CC complete genomic sequencing. Using PCR primers AAQ39419-Q39580 (sequences
 CC designed from the ESTs) sublocalisation of an EST can be achieved with
 CC panels of fragments from specific chromosomes or pools of large genomic
 CC clones in an analogous manner. This sequence represents EST00025.

XX Sequence 385 BP; 101 A; 66 C; 78 G; 139 T; 1 other;

Query Match 56.0%; Score 14; DB 14; Length 385;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 12 tcacagaattattt 25
 |||||
 Db 116 tcacagaattattt 129

RESULT 14

AA059061
 ID AA059061 standard; cDNA; 385 BP.
 XX
 AC AA059061;

DT 16-MAR-1994 (first entry)
 XX
 DE Human brain Expressed Sequence Tag EST00025.
 XX

KW Gene transcription product; genetic markers; tagging; in vivo;
 KW transcription; mapping; locations; chromosomes; chromosomal; ss.
 XX
 OS Homo sapiens.

PN W09316178-A.
 XX

PD 19-AUG-1993.
 XX

PF 12-FEB-1993; 93WO-US01294.
 XX

PR 12-FEB-1992; 92US-0837195.
 XX

PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
 XX

PI Adams MD, Moreno RF, Venter CJ;
 XX

DR WPI: 1993-272882/34.
 XX

PT Enriched oligonucleotides and corresp. sequences - used as
 PT markers for human genes transcribed in-vivo, facilitate tagging
 PT of most human genes
 XX

PS Example 1; Page 115; 500pp; English.
 XX

CC The Expressed Sequence Tag was isolated from a human brain cDNA
 CC library as part of a large set of ESTs which can be used as markers
 CC for human genes transcribed in vivo. They can be used to facilitate
 CC tagging of most human genes, for mapping locations of expressed genes
 CC on chromosomes, for individual or forensic identification, for mapping
 CC locations of disease-associated genes, for identification of tissue
 CC type, and for prepn. of antisense sequences, probes and constructs.
 CC EST00025 has a "poor" coding probability as evaluated using the
 CC coding-region prediction program CRM and hybridises to chromosome 20.
 CC See also AA059041-Q61440.
 XX
 SQ Sequence 385 BP; 99 A; 66 C; 78 G; 140 T; 2 other;

Query Match 56.0%; Score 14; DB 14; Length 385;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 tcacagaattattt 25
 |||||
 Db 116 tcacagaattattt 129

RESULT 15

AAZ56756/c
 ID AAZ56756 standard; cDNA; 902 BP.
 XX
 AC AAZ56756;

XX

DT 23-MAR-2000 (first entry)
 XX

DE Human transmembrane protein HTPMN-59 encoding cDNA.
 XX

KW Human: transmembrane protein; HTPMN; diagnosis; immunospecific;
 KW antiproliferative; neuroprotective; immune disorder;
 KW reproductive disorder; smooth muscle disorder; neurological disorder;
 KW gastrointestinal disorder; developmental disorder;
 KW cell proliferative disorder; ss.
 XX

OS Homo sapiens.
 XX

PN W09961471-A2.
 XX

PD 02-DEC-1999.
 XX

PE 28-MAY-1999; 99WO-US11904.
 XX

PR 29-MAY-1998; 98US-0087260.
 XX

PR 02-JUL-1998; 98US-0091674.
 XX

PR 02-OCT-1998; 98US-0102954.
 XX

PR 24-NOV-1998; 98US-0109869.
 XX

PA (INCY-) INCYTE PHARM INC.
 XX

PI Tang YT, Lal P, Hillman JL, Yue H, Guegler KJ, Corley NC;
 PI Bandman O, Patterson C, Gorgone GA, Kaser MR, Baughn MR;
 PI Au-Young J;
 XX

DR WPI: 2000-072605/06.
 XX

DR P-PSDB; AAY57935.
 XX

PT Proteins, polynucleotides, vectors, host cells and antibodies used to
 PT diagnose, treat or prevent immune, reproductive, smooth muscle,
 PT neurological, gastrointestinal, developmental and cell proliferative
 PT disorders -
 XX

PS Claim 9; Page 215; 229pp; English.
 XX

CC AAZ56698 to AAY56776 encode AAY57877 to AAY57955 which represent human
 CC transmembrane proteins designated HTPMN-1 to HTPMN-79, respectively.
 CC The transmembrane protein have immunospecific, antiproliferative and
 CC neuroprotective activities. The human transmembrane proteins,
 CC polynucleotides encoding them and other compositions and methods from
 CC the present invention, can be used for the diagnosis, treatment or
 CC prevention of immune, reproductive, smooth muscle, neurological,
 CC gastrointestinal, developmental and cell proliferative disorders. The
 CC HTPMN's can be used to treat or prevent disorders associated with a
 CC decreased expression or activity of HTPMN.
 XX
 SQ Sequence 902 BP; 261 A; 212 C; 196 G; 233 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 902;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 agtcacagaattat 23
 |||||
 Db 425 AGTCACAGATTAT 412

RESULT 16

AAAN80927/c
 ID AAAN80927 standard; DNA; 1050 BP.
 XX
 AC AAAN80927;

DT 15-OCT-1990 (first entry)
 XX

DE Lividans Exported Protein gene.
 XX

XX

KW Longisporus trypsin inhibitor; Lividans exported protein; LTI; LEP-10;

KW protease inhibitor; probe: ss.
 XX
 OS Streptomyces lividans strain 1326.
 XX
 FH Key Location/Qualifiers
 FT sig_peptide 394..498
 FT /*tag= a
 FT mat_peptide 499..825
 FT /*tag= b
 FT /product=LEP-10
 FT
 XX
 PN WO8801278-A.
 XX
 PD 25-FEB-1988.
 XX
 PF 17-AUG-1987; 87WO-US02009.
 XX
 PR 18-AUG-1986; 86US-0897245.
 XX
 PA (SMIK) SMITHKLINE BECKMAN CORP.
 PA (BERK/) TR BERKA.
 XX
 PI Berka TR, Fornwald JA, Gorniak JG, Rosenberg M;
 XX
 DR WPI: 1988-063996/09.
 DR P-PSDB; AAR80944.
 XX
 FT Novel protease inhibitors LEP-10 and LTI - produced by Streptomyces spp.
 XX
 PS Disclosure; : PP; English.
 XX
 CC The sequence can be cloned and used to produce recombinant LEP-10.
 CC The DNA can be used as a probe for identifying further coding
 CC sequences for protease inhibitors; the protein is useful as an
 CC assay reagent.
 CC See also AAN80926.
 CC
 SQ Sequence 1050 BP; 149 A; 398 C; 366 G; 137 T; 0 other;

Query Match 56.0%; Score 14; DB 9; Length 1050;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 aagtcacagaatta 22
 |||
 DB 251 AAGTCACAGAAATTA 238

RESULT 17
 AAT31932/C
 ID AAT31932 standard; DNA; 1677 BP.
 XX
 AC AAT31932;
 XX
 DT 27-SEP-1996 (first entry)
 XX
 DE Retinoid X receptor interacting protein R1P14-2 DNA.
 XX
 KW Retinoid X receptor interacting protein; RXR; R1P; R1P14-2; ss.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1356
 FT /*tag= a
 FT 1672-1677
 FT polyA_signal /*tag= b
 XX
 PN WO9621677-A1.
 XX
 PD 18-JUL-1996.

PF 08-DEC-1995; 95WO-US16311.
 XX
 PR 13-JAN-1995; 95US-0372652.
 XX
 PA (GEHO) GEN HOSPITAL CORP.
 XX
 PI Choi H, Moore D, Seol W;
 XX
 DR WPI: 1996-342241/34.
 DR P-PSDB; AAR9739.
 XX
 FT Retinoid X receptor (RXR) interacting protein (RIP) - useful to
 FT modulate or mediate RXR function, anti-RIP antibodies can be used to
 FT determine RIP subcellular distribution patterns
 XX
 PS Claim 8; Page 59-60; 90pp; English.
 XX
 CC A cDNA clone (AAT31932) codes for retinoid X receptor interacting
 CC protein R1P14-2 (AAR9739), a previously undescribed orphan member
 CC of the nuclear receptor superfamily which may be involved in the
 CC complex retinoid response. It was isolated from a mouse liver
 CC cDNA library using an R1P14 probe. Different R1P14-2 clones had
 CC different 5' sequences (see also AAT31935-37). Another isoform,
 CC R1P14-1, was also detected (see also AAT31928). The cDNA can be
 CC used for the recombinant prodn. of R1P14-2 in transformed host
 CC cells.
 XX
 SQ Sequence 1677 BP; 486 A; 385 C; 408 G; 398 T; 0 other;

Query Match 56.0%; Score 14; DB 17; Length 1677;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ttgtcgcaagtcac 15
 |||
 DB 630 TTGTGCGCAAGTCAC 617

RESULT 18
 AAT31928/C
 ID AAT31928 standard; DNA; 1787 BP.
 XX
 AC AAT31928;
 XX
 DT 27-SEP-1996 (first entry)
 XX
 DE Retinoid X receptor interacting protein R1P14-1 DNA.
 XX
 KW Retinoid X receptor interacting protein; RXR; R1P; R1P14-1; ss.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 12..1466
 FT /*tag= a
 FT 1782..1787
 FT polyA_signal /*tag= b
 XX
 PN WO9621677-A1.
 XX
 PD 18-JUL-1996.
 XX
 PF 08-DEC-1995; 95WO-US16311.
 XX
 PR 13-JAN-1995; 95US-0372652.
 XX
 PA (GEHO) GEN HOSPITAL CORP.
 XX
 PI Choi H, Moore D, Seol W;
 XX
 DR WPI: 1996-342241/34.
 DR P-PSDB; AAR9735.

XX Retinoid X receptor (RXR) interacting protein (RIP) - useful to
 PT modulate or mediate RXR function, anti-RIP antibodies can be used to
 PT determine RIP subcellular distribution patterns
 PT
 XX
 XX Claim 8; Page 53-54; 90pp; English.

CC A full-length cDNA clone (AA731928) codes for retinoid X receptor
 CC interacting protein RIP14-1 (AA99735), a previously undescribed
 CC orphan member of the nuclear receptor superfamily which may be
 CC involved in the complex retinoid response. It was isolated from
 CC a mouse liver cDNA library using an RIP14 probe. Another isoform,
 CC RIP14-2, was also detected (see also AA731932). The cDNA can be used
 CC for the recombinant prodn. of RIP14-1 in transformed host cells.
 CC
 SQ Sequence 1787 BP; 515 A; 408 C; 442 G; 422 T; 0 other;

Query Match 56.0%; Score 14; DB 17; Length 1787;
 Best Local Similarity 100.0%; Pred. No. 26;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 ttgtcgcaagtcac 15
 |||||
 Db 740 TTGTCCGCAAGTCAC 727

RESULT 19

AA52940/C
 ID AA52940 standard; cDNA: 1890 BP.

AC AA52940;

DT 14-MAR-2000 (first entry)

DE Human prostate tumor cDNA library derived EST fragment #83.

KW Pancreas; tumor; EST: expressed sequence tag; human; cytostatic;
 KW treatment; ds.

OS Homo sapiens.

PN DE19620190-A1.

PD 04-NOV-1999.

PF 28-APR-1998; 98DE-1020190.

PR 28-APR-1998; 98DE-1020190.

PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

DR WPI: 1999-621386/54.

P-PSDB: AAT74061, AAT74062, AAT74063.

PT New human nucleic acid sequences from pancreatic tumors, and related
 PT proteins -

PS Claim 2; Page 250; 502pp; German.

CC This invention describes novel polypeptides and their encoding nucleic
 CC acids derived from human pancreatic tumor tissue which have cytostatic
 CC activity. The sequences are also useful in producing pharmaceutical
 CC compositions for treatment of pancreatic tumors. AA52858-253014
 CC represent expressed sequence tag (EST) fragments derived from a human
 CC pancreatic tumor cDNA library and which encode the proteins represented
 CC in AAT73814-Y74252.

SQ Sequence 1890 BP; 590 A; 358 C; 398 G; 544 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 1890;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 caagtcacgaatt 21
 |||||
 Db 1475 CAAGTCACGAAATT 1462

RESULT 20

AAV10120/C
 ID AAV10120 standard; cDNA: 2233 BP.

AC AAV10120;

DT 29-MAY-1998 (first entry)

DE Human retinoid receptor RRI cDNA.

KW Retinoid receptor; RRI; steroid receptor; agonist; antagonist; cancer;
 KW adrenal deficiency; skin disorder; inflammatory disorder;

KW Immune response regulator; autoimmune disease; therapeutic antibody; ds.

OS Homo sapiens.

FT Key Location/Qualifiers

FT CDS 363..1781

FT /*Tag= a

FT /product= RRI

FT /note= "retinoid receptor"

PN US5728548-A.

PD 17-MAR-1998.

PF 29-JUN-1995; 95US-0496631.

PR 29-JUN-1995; 95US-0496631.

PA (GENY) GENETICS INST INC.

PI Bowman M;

DR WPI: 1998-206567/18.

P-PSDB: AAM40072.

PT Human retinoid receptor protein RRI - useful for, e.g. drug
 PT screening, therapy and antibody production

PS Claim 8; Column 11-15; 13pp; English.

CC This sequence encodes a novel human steroid receptor, the retinoid
 CC receptor protein or RRI. This protein can be used in screening assays
 CC for steroid hormone receptor agonists and antagonists and in
 CC pharmaceutical compositions for treating adrenal deficiencies, e.g.
 CC Addison's disease, cancer, skin disorders, e.g. acne and psoriasis,
 CC inflammatory disorders, e.g. arthritis and HIV infections. The protein
 CC can also be used for regulating immune responses, e.g. as antitumor
 CC agents, vaccine adjuvants, organ rejection inhibitors or agents for
 CC treating autoimmune diseases. The protein can further be used to produce
 CC therapeutic antibodies.

SQ Sequence 2233 BP; 744 A; 427 C; 463 G; 599 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 2233;
 Best Local Similarity 100.0%; Pred. No. 26;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 ttgtcgcaagtcac 15
 |||||
 Db 1055 TTGTCCGCAAGTCAC 1042

```

RESULT 21
AAAF7541/C
ID   AAF7541 standard; DNA; 2585 BP.
XX
XX   AAF7541;
XX
XX   23-MAY-2001 (first entry)
XX
XX   Caenorhabditis briggsae mab-21 coding sequence.
DE
XX
XX   Mab-21; cell fate choice; pax-6; aniridia; Moebius syndrome;
KW   chromosome 13q13; peripheral neural tissue differentiation; ds.
XX
XX   Caenorhabditis briggsae.
OS
XX
XX   US6197504-B1.
XX
XX   06-MAR-2001.
XX
XX   19-JAN-1998; 98US-0008697.
XX
XX   10-APR-1996; 96US-0011607.
XX   09-APR-1997; 97US-0835604.
XX
XX   (CHOW/) CHOW K L.
XX
XX   Chow KL;
PI
XX   WPI: 2001-256362/26.
XX   P-PSDB; AAB74419.
DR
XX
XX   Detecting expression of mab-21 gene of Caenorhabditis elegans encoding
PT   novel protein required for choice of alternate cell fates in sample, by
PT   contacting mRNA from sample with nucleic acid hybridizing with mab-21 -
XX
XX   Disclosure: column 41-44; 79pp; English.
XX
XX   The present invention describes a method of detecting mab-21 expression
CC   in a sample involving contacting total mRNA from the sample with a probe
CC   to the mab-21 gene. The sequences of the mab-21 coding sequence and
CC   protein from several species are also given. The human mab-21 gene is
CC   found on chromosome 13q13, a region which has been linked to Moebius
CC   syndrome. The mab-21 protein shows homology to pax-6, mutations in which
CC   are associated with aniridia and which is linked to peripheral neural
CC   tissue differentiation.
XX
XX   Sequence 2585 BP; 751 A; 540 C; 559 G; 688 T; 47 other;
SQ
XX
XX   Query Match          56.0%; Score 14; DB 22; Length 2585;
XX   Best Local Similarity 100.0%; Pred. No. 26;
XX   Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX   QY      8 caagtcacagaatt 21
XX          |||||||||||||
XX   DB     1058 CAACTCACAGATT 1045
XX
XX   RESULT 22
XX   AAC69462/c
XX   ID   AAC69462 standard; cDNA; 2845 BP.
XX
XX   AAC69462;
XX
XX   30-JAN-2001 (first entry)
XX
XX   Human secreted protein gene 8 SEQ ID NO:18.
XX
XX   Human: secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW   antithematic; antiproliferative; cytostatic; cardiant; vasotrophic;
KW   cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
KW   fungicide; ophthalmological; gene therapy; autoimmune disease; infection;

```

```

KW   hyperproliferative disorder; cardiovascular disorder; angiogenesis;
KW   cerebrovascular disorder; nervous system disorder; ocular disorder;
KW   wound healing; skin aging; food additive; preservative; ss.
XX
XX   Homo sapiens.
OS
XX   WO200058469-A1.
XX
XX   05-OCT-2000.
XX
XX   23-MAR-2000; 2000WO-US07579.
XX
XX   26-MAR-1999; 99US-0126509.
XX   07-JAN-2000; 2000US-0174853.
XX
XX   (HUMA-) HUMAN GENOME SCI INC.
XX
XX   Rosen CA, Ruben SM, Komatsoulis G;
XX
XX   WPI: 2000-594642/56.
XX   P-PSDB; AAB38210.
XX
XX   Isolated nucleic acid molecule encoding a human secreted protein is
PT   used in preventing, treating or ameliorating a medical condition -
XX
XX   Claim 1; Page 334-335; 416pp; English.
XX
XX   The polynucleotide sequences given in AAC69455 to AAC69502 encode the
CC   human secreted proteins given in AAB38203 to AAB38250. AAB38251 to
CC   AAB38320 represent human secreted polypeptide sequences and proteins
CC   homologous to them, which are given in the exemplification of the present
CC   invention. Human secreted proteins have activities based on the tissues
CC   and cells the genes are expressed in. Example of activities include:
CC   immunosuppressive; antiarthritic; antineumatic; antiproliferative;
CC   cytostatic; cardiant; vasotrophic; cerebroprotective; neurotropic;
CC   neuroprotective; antibacterial; virucide; fungicide; and
CC   ophthalmological. The polynucleotides and polypeptides can be used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Disorders which are diagnosed or treated include
CC   autoimmune diseases, hyperproliferative disorders, cardiovascular
CC   disorders, cerebrovascular disorders, angiogenesis, nervous system
CC   disorders, infections caused by bacteria, viruses and fungi and ocular
CC   disorders. The polypeptides can also be used to aid wound healing and
CC   epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC   maintain organs before transplantation, for supporting cell culture of
CC   primary tissues, to regenerate tissues and in chemotaxis. The
CC   polypeptides can also be used as a food additive or preservative to
CC   increase or decrease storage capabilities. AAC69446 to AAC69454 and
CC   AAB38202 represent sequences used in the exemplification of the present
CC   invention.
XX
XX   Sequence 2845 BP; 871 A; 518 C; 439 G; 1007 T; 10 other;
SQ
XX
XX   Query Match          56.0%; Score 14; DB 21; Length 2845;
XX   Best Local Similarity 100.0%; Pred. No. 26;
XX   Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX   OY      12 tcacagaattatt 25
XX          |||||||||||||
XX   DB     1681 TCACAGATTATT 1668
XX
XX   RESULT 23
XX   AAV72330/c
XX   ID   AAV72330 standard; DNA; 3413 BP.
XX
XX   AAV72330;
XX
XX   28-JUL-1999 (first entry)
XX

```



```

OS Homo sapiens.
XX
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
XX 05-APR-1999; 99US-0127728.
XX 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinketsu RA, Leach M;
XX
XX WPI: 2000-602362/57.
XX P-PSDB; AAB41805.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease.
XX
XX Claim 5; Page 2354-2357; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORF open reading frames 1 to 3161. The ORF
XX sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
XX osteoparalytic; anticonvulsant; antirheumatic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX antihypertensive; and antianemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORF-associated disorder. The
XX nucleic acids can be used to express ORF proteins in gene therapy
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antiinflammatory disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 5773 BP; 1691 A; 1125 C; 1269 G; 1687 T; 1 other;
XX
XX
XX Query Match 56.0%; Score 14; DB 21; Length 5773;
XX Best Local Similarity 100.0%; Pred. No. 26;
XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 7 gcaagtcacagaat 20
XX ||||||||||||
XX Db 4611 GCAAGTCACAGAA 4598
XX
XX
XX RESULT 26
XX AAT10105
XX ID AAT10105 standard; DNA; 8041 BP.
XX
XX AAT10105;
XX
XX 13-MAY-1996 (first entry)
XX
XX Adherence conferring plasmid pear.
XX
XX adhesin: plasmid pear; vector: vaccine; intestine colonisation; ds.
XX
XX Chimeric Escherichia coli;
XX
XX

```

```

OS Chimeric synthetic.
XX
XX Key Location/Qualifiers
XX primer_bind complement (2867..2890)
XX /tag= a
XX /note= "primer for E.coli O157:H7 DNA"
XX complement (2895..2914)
XX
XX primer_bind
XX /tag= b
XX /note= "primer for pear"
XX 3036..5126
XX /tag= c
XX /product= adhesin
XX /trans_except_pos:3912..3914;_aa:GLY
XX /note= "claim 1, page 36"
XX 3271..3310
XX /tag= d
XX /note= "region of transposon TnpH0A insertion"
XX 3801..3840
XX /tag= e
XX /note= "region of transposon TnpH0A insertion"
XX complement (5176..5196)
XX /tag= f
XX /note= "primer for pear"
XX complement (5159..5183)
XX /tag= g
XX /note= "primer for E.coli O157:H7 DNA"
XX complement (6449..7024)
XX /tag= h
XX /note= "tere homologue"
XX complement (7092..7670)
XX /tag= i
XX /note= "terd homologue"
XX
XX
XX WO9600233-A1.
XX
XX 04-JAN-1996.
XX
XX 07-JUN-1995; 95WO-US06994.
XX
XX 24-JUN-1994; 94US-0265714.
XX
XX (CHIL-) CHILDREN'S HOSPITAL & MEDICAL CENT.
XX (UNIM ) UNIV WASHINGTON.
XX (UNIM ) UNIV WASHINGTON STATE RES FOUND.
XX
XX Besser TE, Bilge SS, Tarr PI, Vary JC;
XX
XX WPI: 1996-068826/07.
XX P-PSDB; AAR75366.
XX
XX Chromosomal DNA from E. coli O157:H7 encoding epithelial adhesin -
XX isolated on plasmid psc (overlap), for use as a vaccine to prevent
XX bacterial colonisation of bovine intestine
XX
XX Disclosure; Page 23-30; 42pp; English.
XX
XX Adherence conferring plasmid pear (AAT10105) comprises Escherichia
XX coli O157:H7 chromosomal DNA plus the stragene SK+ vector.
XX It includes 3 open reading frames, 2 of which are homologues of
XX tere and terd genes necessary for tellurite resistance. The
XX third ORF is homologous to the irgA gene and encodes an adhesin
XX (AAR75366) that enables E. coli O157:H7, an antibiotic-resistant,
XX virulent and common food-borne pathogen, to adhere to epithelial
XX cells. This ORF can be utilised in the prodn. of adhesin for use
XX as a vaccine to prevent disease or colonisation of mucosal surfaces
XX by O157:H7.
XX
XX Sequence 8041 BP; 2241 A; 1807 C; 1838 G; 2126 T; 29 other;
XX
XX
XX Query Match 56.0%; Score 14; DB 17; Length 8041;
XX Best Local Similarity 100.0%; Pred. No. 26;
XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX

```

QY 10 agtcacagaattat 23
 |||
 Db 3727 agtcacagaattat 3740

RESULT 27
 AAF21307
 ID AAF21307 standard; DNA: 32351 BP.
 XX
 AC AAF21307;
 XX
 DT 14-MAR-2001 (first entry)
 XX
 DE Human low adenosine antisense oligonucleotide related sequence #2874.
 XX
 KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
 KW human; airway disorder; bronchoconstriction; lung inflammation;
 KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
 KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
 KW respiratory obstruction; pulmonary obstruction; impeded respiration;
 KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
 KW cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200062736-A2.
 XX
 PD 26-OCT-2000.
 XX
 PF 24-MAR-2000; 2000WO-US08020.
 XX
 PR 06-APR-1999; 99US-0127958.
 XX
 PA (UYEC-) UNIV EAST CAROLINA.
 XX
 PA (NYCE/) NYCE J W.
 XX
 PI Nyce JW;
 XX
 DR WPI: 2000-679539/66.
 XX
 PT Low adenosine (A) content antisense oligonucleotides which do not
 PT trigger adenosine receptors during metabolism, useful e.g. for treating
 PT cancers and respiratory obstructions -
 XX
 PS Disclosure: Page 1295-1303; 1592pp; English.

The present invention describes low adenosine (A) content antisense oligonucleotides and compositions (I) comprising them. In the antisense oligonucleotides the A is replaced by a 'Universal' or alternative base. (I) can have respiratory, bronchodilator, antiinflammatory, analgesic, immunosuppressive, antiasthmatic, hypotensive and cytostatic activities. The antisense oligonucleotides and (I) can be used to down-regulate the expression and/or activity of target polypeptides associated with lung/respiratory disorders and malignancies, such as stimulating and activating peptide factors and transmitters, transcription factors, immunoglobulins and antibodies, antibody receptors, cytokines and chemokines, endogenous produced specific and non-specific enzymes, binding proteins, adhesion molecules and their receptors, cytokine and chemokine receptors, adenosine receptors, bradykinin receptors, central nervous system (CNS) and peripheral nervous and non-nervous system peptide transmitters, defensins, growth factors, vasoactive peptides and receptors, binding proteins and malignancy associated proteins. The antisense oligonucleotides may be used in this way to treat disorders including respiratory obstruction (especially pulmonary obstruction and/or bronchoconstriction) and/or lung inflammation, allergies and/or surfactant hypoproduction which are associated with a disease or condition selected from pulmonary vasoconstriction, inflammation, allergies, asthma, impeded respiration, respiratory distress syndrome

CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.
 XX
 SO Sequence 32351 BP; 8594 A; 7026 C; 7405 G; 9326 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 32351;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 aagtcacagaatta 22
 |||
 Db 10696 aagtcacagaatta 10709

RESULT 28
 AAA35185
 ID AAA35185 standard; DNA: 32351 BP.
 XX
 AC AAA35185;
 XX
 DT 28-JUL-2000 (first entry)
 XX
 DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:59.
 XX
 KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphorothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200009525-A2.
 XX
 PD 24-FEB-2000.
 XX
 PF 03-AUG-1999; 99WO-US17712.
 XX
 PR 03-AUG-1998; 98US-0095212.
 XX
 PA (UYEC-) UNIV EAST CAROLINA.
 XX
 PA Nyce JW;
 XX
 PI WPI: 2000-205971/18.
 XX
 PT New antisense oligonucleotides useful for treating e.g. pulmonary
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
 PT cancers -
 XX
 PS Disclosure: Page 1212-1219; 1343pp; English.

The present invention describes a new composition comprising an antisense oligonucleotide (ON) with low adenosine (up to 15%), which targets nucleic acids involved in bronchoconstriction, allergies, and/or inflammation. The ON can have antiinflammatory, antiallergic, antiasthmatic, cytostatic and analgesic activities. The compositions are useful for the treatment of diseases associated with inflammation, impaired airways, including lung disease and diseases whose secondary effects afflict the lungs of a subject. They can be used for treating e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma, impeded respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,

CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONS reduces side effects. The A-containing ONS break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA3312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONS from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.

XX Sequence 32351 BP; 8594 A; 7027 C; 7405 G; 9325 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 32351;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 aagtcacagaatta 22
|||||
Db 10696 aagtcacagaatta 10709

RESULT 29

AAF21311
ID AAF21311 standard; DNA; 40298 BP.

XX AAF21311;

DT 14-MAR-2001 (first entry)

DE Human low adenosine antisense oligonucleotide related sequence #2878.

XX Low adenosine antisense oligonucleotide: phosphorothioate; allergy;
KM human: airway disorder; bronchoconstriction; lung inflammation;
KM surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KM immunosuppressive; analgesic; hypotensive; cytostatic;
KM respiratory obstruction; pulmonary obstruction; impeded respiration;
KM surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KM respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KM pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KM chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KM cancer; SS.

XX Homo sapiens.

OS WO200062736-A2.

PN 26-OCT-2000.

XX 24-MAR-2000; 2000WO-US08020.

PF 06-APR-1999; 99US-0127958.

XX (UYEC-) UNIV EAST CAROLINA.

PA (NYCE/) NYCE J W.

PI NYCE JW;

XX WPI; 2000-679539/66.

DR Low adenosine (A) content antisense oligonucleotides which do not
XX trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
XX
PS Disclosure; Page 1305-1315; 1592pp; English.

CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.

CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotide and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.

XX Sequence 40298 BP; 10485 A; 9119 C; 9484 G; 11210 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 40298;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 aagtcacagaatta 22
|||||
Db 11521 aagtcacagaatta 11534

RESULT 30

AAA35189
ID AAA35189 standard; DNA; 40298 BP.

XX AAA35189;

AC 28-JUL-2000 (first entry)

XX Human adenosine receptor related polynucleotide 2nd SEQ ID NO:63.

DE Human; adenosine receptor; low adenosine antisense oligonucleotide;
XX phosphorothioate; impaired respiration; inflammation; allergy;
XX allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KM antiasthmatic; cytostatic; analgesic; impeded airway;
KM lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KM respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KM pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KM cancer; leukaemia; lymphoma; carcinoma; metastasis; SS.

XX Homo sapiens.

OS WO200009525-A2.

PN 24-FEB-2000.

XX 03-AUG-1999; 99WO-US17712.

XX 03-AUG-1998; 98US-0095212.

XX (UYEC-) UNIV EAST CAROLINA.

PA NYCE JW;

PI NYCE JW;

XX WPI; 2000-205971/18.


```

FT      /*tag= aa
FT      /note= "primer 20-853-.pu"
FT      complement (77166..77185)
FT      /*tag= ab
FT      /note= "primer 20-842..rp"
FT      1220..1238
FT      /*tag= ac
FT      /note= "primer 20-828-311.mis"
FT      complement (1240..1258)
FT      /*tag= ad
FT      /note= "primer 20-828-311.mis"
FT      12328..12346
FT      /*tag= ae
FT      /note= "primer 17-42-319.mis"
FT      complement (12348..12366)
FT      /*tag= af
FT      /note= "primer 17-42-319.mis"
FT      15222..15240
FT      /*tag= ag
FT      /note= "primer 17-41-250.mis"
FT      complement (15242..15260)
FT      /*tag= ah
FT      /note= "primer 17-41-250.mis"
FT      42199..42237
FT      /*tag= ai
FT      /note= "primer 20-841-149.mis"
FT      complement (42219..42237)
FT      /*tag= aj
FT      /note= "primer 20-841-149.mis"
FT      45423..45441
FT      /*tag= ak
FT      /note= "primer 20-842-115.mis"
FT      complement (45443..45461)
FT      /*tag= al
FT      /note= "primer 20-842-115.mis"
FT      77039..77057
FT      /*tag= am
FT      /note= "primer 20-853-415.mis"
FT      complement (77059..77077)
FT      /*tag= an
FT      /note= "primer 20-853-415.mis"
FT      1227..1251
FT      /*tag= ao
FT      /note= "20-828-311 probe"
FT      12335..12359
FT      /*tag= ap
FT      /note= "17-42-319 probe"
FT      15229..15253
FT      /*tag= aq
FT      /note= "17-41-250 probe"
FT      42206..42230
FT      /*tag= ar
FT      /note= "20-841-149 probe"
FT      45430..45454
FT      /*tag= as
FT      /note= "20-842-115 probe"
FT      77046..77070
FT      /*tag= at
FT      /note= "20-853-415 probe"
XX
XX      WO200100803-A2.
XX
XX      04-JAN-2001.
XX
XX      21-JUN-2000; 2000WO-1B01011.
XX
XX      25-JUN-1999; 99US-0141032.
XX      20-DEC-1999; 99WO-1B02058.
XX      21-DEC-1999; 99US-0469099.
XX
XX      (GEST ) GENSET.
XX
XX      Yen F, Denison B, Bour B, Bihain B, Bougueleret L, Duclert A;
PI

```

```

PI      Dumas Milne Edwards J;
XX
XX      MPI: 2001-071485/08.
DR      P-PSDB; AAB20103.
XX
XX      Nucleic acids encoding apolipoprotein A-IV-related proteins (AA4RP) and
PT      diallelic markers of AA4RP, useful for diagnosing lipid metabolism
PT      related disorders and/or liver related disorders.
XX
XX      Claim 1; Page 216-238; 260pp; English.
XX
XX      The present sequence is that of the human apolipoprotein
CC      A-IV-related protein (AA4RP) gene on chromosome 11. The gene
CC      encodes a 366-amino acid protein (see AAB20103). AA4RP is
CC      differentially expressed in obese mouse models, indicating a role
CC      in lipid metabolism related disorders. It appears to be the human
CC      homologue of rat regeneration associated protein (RAP3), which is
CC      believed to be involved in liver regeneration, and is likely to
CC      have a similar function. It also shows 52% similarity to
CC      apolipoprotein A-IV, and is likely to have a similar function. The
CC      invention also provides AA4RP cDNAs (see AAF30036) and polypeptides,
CC      diallelic markers identified in the AA4RP gene and from genomic
CC      regions flanking the gene, and methods for genotyping a nucleic
CC      acid containing 1 or more of the diallelic markers. Also provided
CC      is a method for detecting a statistical correlation between a
CC      diallelic marker allele and a phenotype and/or between a diallelic
CC      marker haplotype and a phenotype. Diagnostic methods are provided
CC      for determining whether an individual is at risk of developing a
CC      lipid metabolism related disorder and/or a liver related disorder,

```

```

Query Match          56.0%; Score 14; DB 22; Length 81001;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      11 gtcacagaattatt 24
      |||
Db      58568 gtcacagaattatt 58581

```

```

RESULT 32
AAZ01425
ID      AAZ01425 standard; DNA; 1038602 BP.
XX
XX      AAZ01425;
AC      XX
XX      07-OCT-1999 (first entry)
XX
XX      Complete genome sequence of Chlamydia trachomatis.
DE      XX
XX      Vaccines: eye disease; conventional trachoma; nonendemic trachoma;
KW      paratrachoma; inclusion conjunctivitis; genital disease; peritrichitis;
KW      nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KW      Bartholinitis; pneumonia; venereal lymphogranulomatosis; ss.
XX
XX      Chlamydia trachomatis.
XX
XX      WO9928475-A2.
XX
XX      10-JUN-1999.
XX
XX      27-NOV-1998; 98WO-1B01939.
XX
XX      04-NOV-1998; 98US-0107077.
XX      28-NOV-1997; 97FR-0015041.
XX      17-DEC-1997; 97FR-0016034.
XX
XX      (GEST ) GENSET.
XX
XX      Grifffals R;
XX
XX      MPI; 1999-371125/31.
XX
XX      Genome sequence of Chlamydia trachomatis
PT

```

XX Claim 1: Page 373-656; 1755pp; English.

PS

XX The present sequence represents the complete genome of Chlamydia

CC trachomatis. Open reading frames (ORFs) of the genome encode

CC polypeptides AAY6754-Y37949. The polypeptides can be used as vaccines

CC against Chlamydia trachomatis. Antisense and ribozyme sequences can also

CC be used to control growth of the microorganism. Chlamydia trachomatis is

CC responsible for a large number of diseases, e.g. eye diseases such as

CC conventional trachoma, nongonococcal urethritis, paratrachoma, and inclusion

CC conjunctivitis; genital diseases such as nongonococcal urethritis,

CC epididymitis, cervicitis, salpingitis, perihepatitis, bartolinitis;

CC pneumopathy in breast feeding infants; and venereal

CC lymphogranulomatosis. The polypeptides of the invention may be of use in

CC treating these diseases.

XX

SQ Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

Query Match 56.0%; Score 14; DB 20; Length 1038602;

Best Local Similarity 100.0%; Pred. NO. 25;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 attgcgcagatca 14

|||||

Db 736214 attgcgcagatca 736227

RESULT 33

AAT22576

ID AAT22576 standard; cDNA to mRNA; 87 BP.

XX

AC AAT22576;

XX

DT 01-OCT-1996 (first entry)

XX

DE Human gene signature H0MGS04192.

XX

KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;

KW human; cloning; mapping; non-biased library; diagnosis; detection;

KW cell typing; abnormal cell function; ss.

XX

OS Homo sapiens.

XX

PN WO9514772-A1.

XX

PD 01-JUN-1995.

XX

PF 11-NOV-1994; 94WO-JP01916.

XX

PR 12-NOV-1993; 93JP-0355504.

XX

PA (MATS/) MATSUBARA K.

PA (OKUBO/) OKUBO K.

XX

PI Matsubara K, Okubo K;

XX

DR WPI: 1995-206931/27.

XX

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.

PT for diagnosis of abnormal cell function, by preparing cDNA that

PT reflects relative abundance of corresp. mRNA in specific human

PT tissues

XX

PS Claim 1: Page 1160; 2245pp; Japanese.

XX

CC A single-stranded DNA (or its complementary strand or the corresp.

CC double-stranded DNA) which comprises one of the 7837 "GS" sequences

CC given in AAT19001-T26837 and which is able to hybridise to part of

CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)

CC sequences were obtained from 3'-directed cDNA libraries prepared

CC from various human tissues; synthesis of cDNA was initiated from the

CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

CC untranslated sequence is unique to a particular mRNA species, almost

CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library

CC is constructed so as to reflect accurately the relative abundance of

CC different mRNAs in the particular tissue from which it was derived.

CC The appearance frequency of a given GS in a cDNA library can be

CC determined (esp. using primers and probes derived from the GS

CC sequences) as a means of diagnosing abnormal cell function or for

CC recognising different cell types.

XX

SQ Sequence 87 BP; 31 A; 25 C; 9 G; 21 T; 1 other;

Query Match 52.0%; Score 13; DB 16; Length 87;

Best Local Similarity 100.0%; Pred. NO. 98;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 aagtcacagaatt 21

|||||

Db 16 aagtcacagaatt 28

RESULT 34

AAT20380

ID AAT20380 standard; cDNA to mRNA; 105 BP.

XX

AC AAT20380;

XX

DT 19-JUL-1996 (first entry)

XX

DE Human gene signature H0MGS051532.

XX

KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;

KW human; cloning; mapping; non-biased library; diagnosis; detection;

KW cell typing; abnormal cell function; ss.

XX

OS Homo sapiens.

XX

PN WO9514772-A1.

XX

PD 01-JUN-1995.

XX

PF 11-NOV-1994; 94WO-JP01916.

XX

PR 12-NOV-1993; 93JP-0355504.

XX

PA (MATS/) MATSUBARA K.

PA (OKUBO/) OKUBO K.

XX

PI Matsubara K, Okubo K;

XX

DR WPI: 1995-206931/27.

XX

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.

PT for diagnosis of abnormal cell function, by preparing cDNA that

PT reflects relative abundance of corresp. mRNA in specific human

PT tissues

XX

PS Claim 1: Page 625; 2245pp; Japanese.

XX

CC A single-stranded DNA (or its complementary strand or the corresp.

CC double-stranded DNA) which comprises one of the 7837 "GS" sequences

CC given in AAT19001-T26837 and which is able to hybridise to part of

CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)

CC sequences were obtained from 3'-directed cDNA libraries prepared

CC from various human tissues; synthesis of cDNA was initiated from the

CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

CC untranslated sequence is unique to a particular mRNA species, almost

CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library

CC is constructed so as to reflect accurately the relative abundance of

CC different mRNAs in the particular tissue from which it was derived.

CC The appearance frequency of a given GS in a cDNA library can be

CC determined (esp. using primers and probes derived from the GS

CC sequences) as a means of diagnosing abnormal cell function or for

CC recognising different cell types.
XX
Sequence 105 BP; 26 A; 19 C; 20 G; 34 T; 6 other;

Query Match
Best Local Similarity 52.0%; Score 13; DB 16; Length 105;
Matches 13; Conservative 100.0%; Pred. No. 98;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 cacagattatt 25
|||||
DB 92 cacagattatt 104

RESULT 35

AAT20233/c
ID AAT20233 standard; cDNA to mRNA; 167 BP.

AC AAT20233;

DT 24-JUL-1996 (first entry)

DE Human gene signature HUNGSO1380.

XX Gene signature: messenger RNA; mRNA; relative abundance; frequency;

KW human; cloning; mapping; non-biased library; diagnosis; detection;

XX cell typing; abnormal cell function; ss.

OS Homo sapiens.

XX W09514772-A1.

XX 01-JUN-1995.

XX 11-NOV-1994; 94WO-0P01916.

XX 12-NOV-1993; 93JP-0355504.

XX (MATS/) MATSUBARA K.

XX (OKUBO/) OKUBO K.

XX Matsubara K, Okubo K;

XX WPI: 1995-206931/27.

XX Identifying gene signatures in 3'-directed human cDNA library - e.g.

XX for diagnosis of abnormal cell function, by preparing cDNA that

XX reflects relative abundance of corresp. mRNA in specific human

XX tissues

XX Claim 1: Page 592; 2245pp; Japanese.

XX A single-stranded DNA (or its complementary strand or the corresp.

XX double-stranded DNA) which comprises one of the 7837 "GS" sequences

XX given in AAT19001-T26837 and which is able to hybridise to part of

XX human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)

XX sequences were obtained from 3'-directed cDNA libraries prepared

XX from various human tissues; synthesis of cDNA was initiated from the

XX 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

XX untranslated sequence is unique to a particular mRNA species, almost

XX all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library

XX is constructed so as to reflect accurately the relative abundance of

XX different mRNAs in the particular tissue from which it was derived.

XX The appearance frequency of a given GS in a cDNA library can be

XX determined (esp. using primers and probes derived from the GS

XX sequences) as a means of diagnosing abnormal cell function or for

XX recognising different cell types.

XX Sequence 167 BP; 65 A; 23 C; 26 G; 44 T; 9 other;

XX Query Match
Best Local Similarity 52.0%; Score 13; DB 16; Length 167;
Matches 13; Conservative 100.0%; Pred. No. 97;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 agtcacagaatta 22
|||||

DB 118 AGTCACAGAAATTA 106

RESULT 36

AAC26213/c
ID AAC26213 standard; cDNA; 187 BP.

AC AAC26213;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 30288.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

XX gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclet A, Giordano J;

XX WPI: 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for

XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

XX diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1: SEQ ID 30288; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from

XX mRNAs encoding secreted proteins. No ORF has yet been conclusively

XX identified within the present sequence. The 5' ESTs were prepared from

XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST

XX sequences usually correspond mainly to the 3' untranslated region (UTR)

XX of the mRNA because they are often obtained from oligo-dT primed cDNA

XX libraries. Such ESTs are not well suited for isolating cDNA sequences

XX derived from the 5' ends of mRNAs and even in those cases where longer

XX cDNA sequences have been obtained, the full 5' UTR is rarely included.

XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be

XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used

XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX They are used to obtain upstream regulatory sequences and to design

XX expression and secretion vectors.

XX Sequence 187 BP; 59 A; 27 C; 30 G; 71 T; 0 other;

XX Query Match
Best Local Similarity 52.0%; Score 13; DB 21; Length 187;
Matches 13; Conservative 100.0%; Pred. No. 97;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX QY 9 aagtcacagaatt 21
|||||

XX DB 163 AAGTCACAGAAATT 151

XX RESULT 37
AAO77082
ID AAO77082 standard; DNA; 354 BP.

```

AC  AAQ77082;
XX
XX  23-SEP-1994 (first entry)
XX
XX  Human genome fragment.
DE
XX  Brain; placenta; bone marrow; genetic analysis; gene mapping;
XX  detection; homology; human; adrenal tissue; ds.
XX
XX  Homo sapiens.
XX
XX  WO9401548-A.
XX
XX  20-JAN-1994.
XX
XX  13-JUL-1993; 93WO-GB01467.
XX
XX  13-JUL-1992; 92GB-0014857.
XX
XX  (MEDI-) MEDICAL RES COUNCIL.
XX
XX  Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;
XX  Sibson DR, Starkey M;
XX  WPI; 1994-035056/04.
XX
XX  New nucleic acid fragment encoding gene products - can be used
XX  for genetic analysis and mapping
XX
XX  Claim 1: Page 347; 616pp; English.
XX
XX  Human nucleic acid fragments, isolated from brain adrenal tissue,
XX  the placenta or bone marrow comprise any of: (A) a sequence
XX  selected from (AAQ76401-Q77613), (B) an allelic variation of a
XX  sequence as described in (A), or (C) a sequence complementary
XX  to (A) or (B).
XX
XX  Sequence 354 BP; 92 A; 92 C; 86 G; 84 T; 0 other;
XX
XX
XX  Query Match          52.0%; Score 13; DB 15; Length 354;
XX  Best Local Similarity 100.0%; Pred. No. 97;
XX  Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  Qy      12 tcacagaattatt 24
XX          |||||
XX  Db      192 tcacagaattatt 204
XX
XX  RESULT 38
XX  AAA50159/C
XX  ID      AAA50159 standard; cDNA: 363 BP.
XX
XX  AC      AAA50159;
XX
XX  DT      07-NOV-2000 (first entry)
XX
XX  DE      Antibody 5H7 heavy chain variable region cDNA.
XX
XX  KW      5H7: monoclonal antibody; apoptosis; programmed cell death;
XX  major histocompatibility complex class I; MHC; tumour; therapy;
XX  graft-mediated immunoprotection; single chain antibody; scfv; ss.
XX
XX  OS      Mammalia.
XX
XX  PN      WO200047713-A2.
XX
XX  PD      17-AUG-2000.
XX
XX  PF      08-FEB-2000; 2000WO-US03234.
XX
XX  PR      09-FEB-1999; 99US-0119238.
XX

```

```

PA  (ARCH-) ARCH DEV CORP.
XX
XX  Woodle ES, Van Severter JM, Kulkarni S, Kranz D, Holman P;
XX
XX  WPI; 2000-558210/51.
XX
XX  P-PSDB; AAY95776.
XX
XX  New polynucleotides encoding single chain light and heavy variable
XX  regions of a 5H7 antibody, useful in methods for conferring programmed
XX  cell death and graft-mediated immune protection of cell, tissue and
XX  organs -
XX
XX  Claim 2; Page 36; 48pp; English.
XX
XX  The present sequence is that of cDNA coding for the heavy chain
XX  variable region (VH) (see AAY95776) of 5H7, an anti-human class I
XX  major histocompatibility complex (MHC) monoclonal antibody (MAb)
XX  which recognises a monomorphic determinant of the alpha-3 domain.
XX  Mab 5H7 has been shown to induce programmed cell death (PCD) in
XX  lymphoid tumours, peripheral blood mononuclear cells, and B6 mouse
XX  splenocytes transgenic for class I HLA-B*27. Polynucleotides
XX  encoding 5H7 VL and VH sequences were used in the construction of
XX  single chain variable immunoglobulin domains (scfv) of the 5H7
XX  antibody (see AAY95781). These can be used to confer PCD properties
XX  to cells, especially T and B lymphocyte tumour cells. Recombinant
XX  DNA molecules encoding 5H7 scfv are also used in methods for
XX  conferring graft-mediated immune protection of cell, tissue and
XX  transplanted organs.
XX
XX  Sequence 363 BP; 88 A; 101 C; 84 G; 90 T; 0 other;
XX
XX
XX  Query Match          52.0%; Score 13; DB 21; Length 363;
XX  Best Local Similarity 100.0%; Pred. No. 97;
XX  Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  Qy      9 aagtcacagaatt 21
XX          |||||
XX  Db      262 AAGTCACAGAAATT 250
XX
XX  RESULT 39
XX  AAQ25667/C
XX  ID      AAQ25667 standard; cDNA: 378 BP.
XX
XX  AC      AAQ25667;
XX
XX  DT      28-DEC-1992 (first entry)
XX
XX  DE      Sequence of the anti-urokinase antibody kappa variable region (VK)
XX  cDNA.
XX
XX  KW      Chimeric monoclonal antibody; anti-urokinase antibody; PCR;
XX  antithrombotic agent; myocardial infarction therapy; ss.
XX
XX  OS      Mus musculus.
XX
XX  FH      Key
XX  FT      sig-peptide      Location/Qualifiers
XX  FT      mat_peptide      1..9
XX  FT      misc_feature      10..378
XX  FT      misc_feature      1..27
XX  FT      misc_feature      /tag= C
XX  FT      misc_feature      /label= VH01 primer
XX  FT      misc_feature      18..32
XX  FT      misc_feature      /tag= d
XX  FT      misc_feature      /label= 5'VH1 primer
XX  FT      misc_feature      340..371
XX  FT      misc_feature      /tag= e
XX  FT      misc_feature      /label= 3'mVH primer
XX  FT      misc_feature      356..371
XX  FT      misc_feature      /tag= f
XX

```

FT XX /label= JH01 primer
 PN XX EP491351-A.
 PD XX 24-JUN-1992.
 XX
 PF 17-DEC-1991; 91EP-0121591.
 XX
 PR 18-DEC-1990; 90JP-0413829.
 XX 11-NOV-1991; 91JP-0294464.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 PI Iwasa S, Taka H, Watanabe T, Tada H;
 DR WPI: 1992-209528/26.
 DR P-PSDB; AAR24722.
 XX
 PT Chimeric monoclonal antibodies - contain anti-human fibrin
 PT antibody light and heavy chain variable and constant for treating
 PT thrombotic conditions e.g. myocardial infarction
 XX
 PS Example; Figure 15; 87pp; English.
 XX
 CC Poly(A)+ RNA was prep'd. from mouse anti-uropkinase
 CC antibody-producing hybridoma UK1-3 cells. Using this poly(A) RNA as
 CC a template, an anti-uropkinase antibody VK cDNA was amplified with
 CC the mc-kappa primer as primer for first strand synthesis, and the
 CC 3'mv-kappa and 5'mv-kappa primers as primers for the PCR. The
 CC amplified fragment was restriction digested and ligated into a
 CC restriction fragment of pTB1423 to give an anti-uropkinase antibody
 CC VK cDNA contg. plasmid pTB1456. The sequence of this plasmid is
 CC given in AAQ25667. The cDNA is a functional VK gene. The sequence of
 CC the primer (mc-gamma-1) is given in AAQ25689.
 XX
 SQ Sequence 378 BP; 92 A; 100 C; 91 G; 95 T; 0 other;

Query Match 52.0%; Score 13; DB 13; Length 378;
 Best Local Similarity 100.0%; Pred. No. 97;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 aagtcacagaatt 21
 |||||
 DB 271 AAGTCACAGAAATT 259

RESULT 40
 AAC79285
 ID AAC79285 standard; cDNA; 387 BP.
 XX
 AC AAC79285;
 XX
 DT 05-FEB-2001 (first entry)
 XX
 DE Human lung tumour-specific cDNA #238.
 XX
 KW Lung tumour protein; lung cancer; cytostatic; vaccine; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200060077-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 30-MAR-2000; 2000WO-US08560.
 XX
 PR 02-APR-1999; 99US-0285323.
 PR 09-AUG-1999; 99US-0370838.
 PR 30-DEC-1999; 99US-0476235.
 PR 03-MAR-2000; 2000US-0518809.
 XX
 PA (CORI-) CORIXA CORP.

XX
 PI Reed SG, Lodes MJ, Mohamath R, Secrist H;
 DR WPI: 2000-638466/61.
 XX
 PT Novel lung tumor polypeptides and polynucleotides, useful for
 PT detecting, monitoring or treating cancer, especially lung cancer -
 XX
 PS Claim 24; Page 223; 243pp; English.
 XX
 CC The present sequence is given in a specification relating to compounds
 CC for therapy and diagnosis of lung cancer. Polypeptides comprising at
 CC least an immunogenic part of a lung tumour protein are disclosed.
 CC The polypeptides are useful for inhibiting the development of cancer,
 CC especially lung cancer. Samples of T cells expressing the polypeptides
 CC may be used to inhibit the development of cancer. The polypeptides are
 CC also useful for detecting and monitoring the progression of cancer,
 CC especially lung cancer.
 XX
 SQ Sequence 387 BP; 95 A; 91 C; 99 G; 102 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 387;
 Best Local Similarity 100.0%; Pred. No. 97;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 tcacagaattatt 24
 |||||
 DB 179 tcacagaattatt 191

RESULT 41
 AAA74604/C
 ID AAA74604 standard; cDNA; 408 BP.
 XX
 AC AAA74604;
 XX
 DT 08-JAN-2001 (first entry)
 XX
 DE Murine 5B3 antibody heavy chain variable region cDNA.
 XX
 KW Mouse; 5B3 antibody; IgG1; piezoelectric immunoassay;
 KW small molecule explosive detection; 2,4,6-trinitrotoluene; TNT; ss.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..408
 FT /*tag= a
 FT /partial
 FT /product= "5B3 antibody heavy chain variable region"
 XX
 PN WO200043774-A2.
 XX
 PD 27-JUL-2000.
 XX
 PF 25-JAN-2000; 2000WO-IL00048.
 XX
 PR 25-JAN-1999; 99IL-0128212.
 XX
 PA (YISS) YISSUM RES & DEV CO.
 PA (YEDA) YEDA RES & DEV CO LTD.
 XX
 PI Willner I, Eshhar Z;
 XX
 DR WPI: 2000-524259/47.
 DR P-PSDB; AAB15672.
 XX
 PT Apparatus for detecting small molecules, especially explosives
 PT comprises a piezoelectric crystal -
 XX
 PS Disclosure; Fig 3A; 90pp; English.

CC The present sequence encodes the heavy chain variable region of 5B3
 CC antibody. It was obtained from total RNA extracted from a 5B3 hybridoma
 CC by RT-PCR. 5B3 is an IgG1 antibody derived from a TNP-KLH immunised
 CC mouse. Its binding to TNP or DNS antigen can be blocked by very low
 CC amounts of TNP and it can therefore be used in a method for detecting
 CC small assayed explosive molecules. Molecules are detected using a
 CC piezoelectric sensor. Piezoelectric Immunoassaying in liquid phase
 CC allows stationary and flow analysis of an aqueous sample. The method
 CC is sufficiently sensitive for detection of low molecular weight
 CC molecules.
 CC
 SQ Sequence 408 BP; 90 A; 108 C; 99 G; 111 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 408;
 Best Local Similarity 100.0%; Pred. No. 97;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 aagtcacagaatt 21
 |||||
 DB 316 AAGTCACAGAATT 304

RESULT 42
 AAC03836
 ID AAC03836 standard; CDNA: 409 BP.
 XX
 AC AAC03836;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST. SEQ ID NO: 3834.

KM Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KM gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP103401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI: 2000-500381/45.

DR P-PSDB; AAG03830.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining CDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 Claim 1; SEQ ID 3834; 71pp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.

XX
 SQ Sequence 409 BP; 95 A; 101 C; 118 G; 91 T; 4 other;

Query Match 52.0%; Score 13; DB 21; Length 409;
 Best Local Similarity 100.0%; Pred. No. 97;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 tcacagattatt 24
 |||||
 DB 233 tcacagattatt 245

RESULT 43
 AAC56783/c
 ID AAC56783 standard; DNA: 415 BP.
 XX
 AC AAC56783;

DT 25-JAN-2001 (first entry)

DE Eucalyptus grandis transcription factor DNA sequence #654.

XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
 KM poplar; sweetgum; teak; mahogany; bzlp; G-box binding factor;
 KM basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
 KM homeodomain zipper; LIM domain; AP2; ERBS; zinc finger domain;
 KM type 2 Cys2His2; CCAAT box element; MYB; ss.

OS Eucalyptus grandis.

PN W0200053724-A2.

PD 14-SEP-2000.

PF 09-MAR-2000; 2000WO-US06112.

PR 11-MAR-1999; 99US-0266513.

PR 18-AUG-1999; 99US-0149485.

PA (GENE-) GENESIS RES & DEV CORP LTD.

PI (FLET-) FLETCHER CHALLENGE FORESTS LTD.

PI Wood M, McGrath A, Shenk MA, Glenn M;

DR WPI: 2000-579369/54.

XX New isolated polynucleotide encoding a plant transcription factor for
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
 PT having modified gene expression or modified activity of a polypeptide
 PT -
 XX
 Claim 1; Pages 513-514; 747pp; English.

XX The present invention relates to novel plant transcription factors from
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
 CC sequence for one such transcription factor. The transcription factor may
 CC be used to produce a plant having modified gene expression such as a
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
 CC mahogany species or to modify the activity of a polypeptide in a plant.
 CC The transcription factors of the present invention are members from the
 CC following families of regulatory proteins: bzlp, bzlp family of G-box
 CC binding factors, basic helix-loop-helix zipper,
 CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
 CC and ERBS, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 CC and MYB.

XX
 SQ Sequence 415 BP; 124 A; 99 C; 88 G; 104 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 415;
 Best Local Similarity 100.0%; Pred. No. 97;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 attgtcgaatc 13
 |||||
 DB 121 ATTGTGCAAGTC 109

RESULT 44
 AAC07649/c
 ID AAC07649 standard; cDNA: 441 BP.

AC AAC07649;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 11724.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PS (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

DR New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 11724; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

SO Sequence 441 BP; 123 A; 103 C; 113 G; 102 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 441;
 Best Local Similarity 100.0%; Pred. No. 97;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 cacagaaattattt 25
 |||||
 DB 419 CACAGAAATTATT 407

RESULT 45
 AAC07951
 ID AAC07951 standard; cDNA: 476 BP.
 XX
 AC AAC07951;

XX 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 12026.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PS (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

DR New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 12026; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

SO Sequence 476 BP; 117 A; 113 C; 135 G; 107 T; 4 other;

Query Match 52.0%; Score 13; DB 21; Length 476;
 Best Local Similarity 100.0%; Pred. No. 97;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 tcacagaaattatt 24
 |||||
 DB 300 tcacagaaattatt 312

Search completed: October 9, 2001, 15:52:06
 Job time: 15087 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:55:17 ; Search time 218.82 Seconds
(without alignments)
21.629 Million cell updates/sec

Title: US-09-396-196f-2

Perfect score: 25

Sequence: 1 atgttcgacagtcacagaattatctt 25

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 324599 seqs, 94655562 residues

Word size : 9

Total number of hits satisfying chosen parameters: 8765

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/prodata/2/ina/2A.COMB.seq:*
2: /cgn2_6/prodata/2/ina/2B.COMB.seq:*
3: /cgn2_6/prodata/2/ina/2A.COMB.seq:*
4: /cgn2_6/prodata/2/ina/2B.COMB.seq:*
5: /cgn2_6/prodata/2/ina/2A.COMB.seq:*
6: /cgn2_6/prodata/2/ina/2B.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	25	100.0	1041	2	US-08-401-068-7	Sequence 7, Appl
2	25	100.0	1041	2	US-08-846-338-7	Sequence 7, Appl
3	25	100.0	5872	3	US-08-411-7688-1	Sequence 1, Appl
4	25	100.0	5872	3	US-08-411-7688-6	Sequence 6, Appl
5	17	68.0	17	2	US-07-944-259-2	Sequence 2, Appl
6	15	60.0	1771	2	US-09-166-203-48	Sequence 48, Appl
7	14	56.0	1050	6	5498529-7	Patent No. 5498529
8	14	56.0	1677	5	US-08-372-652-14	Sequence 14, Appl
9	14	56.0	1677	5	PCT-US95-16311-14	Sequence 14, Appl
10	14	56.0	1787	5	US-08-372-652-6	Sequence 6, Appl
11	14	56.0	1787	5	PCT-US95-16311-6	Sequence 6, Appl
12	14	56.0	2091	1	US-08-765-081-4	Sequence 4, Appl
13	14	56.0	2091	1	US-09-098-082-4	Sequence 4, Appl
14	14	56.0	2091	5	PCT-US95-06994-4	Sequence 4, Appl
15	14	56.0	2233	1	US-08-496-631-1	Sequence 1, Appl
16	14	56.0	2585	4	US-09-008-697A-7	Sequence 7, Appl
17	14	56.0	8041	1	US-08-765-081-1	Sequence 1, Appl
18	14	56.0	8041	3	US-09-098-082-1	Sequence 1, Appl
19	14	56.0	8041	3	PCT-US95-06994-1	Sequence 1, Appl
20	14	56.0	1500	4	US-08-875-847B-3	Sequence 3, Appl
21	13	52.0	1505	4	US-08-875-847B-1	Sequence 1, Appl
22	13	52.0	1662	1	US-08-651-572-1	Sequence 1, Appl
23	13	52.0	1662	3	US-09-066-54A-1	Sequence 1, Appl
24	13	52.0	1662	3	US-08-951-086-1	Sequence 1, Appl
25	13	52.0	2214	2	US-08-484-993B-1	Sequence 1, Appl
26	13	52.0	2214	2	US-08-484-158B-1	Sequence 1, Appl
27	13	52.0	2214	2	US-08-484-596A-1	Sequence 1, Appl

c 28	13	52.0	2214	2	US-08-480-150A-1	Sequence 1, Appl
c 29	13	52.0	2214	3	US-08-458-731-1	Sequence 1, Appl
c 30	13	52.0	2214	3	US-08-149-223A-1	Sequence 1, Appl
c 31	13	52.0	2381	2	US-08-484-993B-9	Sequence 9, Appl
c 32	13	52.0	2381	2	US-08-484-158B-9	Sequence 9, Appl
c 33	13	52.0	2381	2	US-08-484-596A-9	Sequence 9, Appl
c 34	13	52.0	2381	3	US-08-480-150A-9	Sequence 9, Appl
c 35	13	52.0	2381	3	US-08-458-731-9	Sequence 9, Appl
c 36	13	52.0	2381	3	US-08-149-223A-9	Sequence 9, Appl
c 37	13	52.0	3083	1	US-07-718-575-5	Sequence 5, Appl
c 38	13	52.0	3083	1	US-08-481-206-5	Sequence 5, Appl
c 39	13	52.0	3083	2	US-08-486-269A-5	Sequence 5, Appl
c 40	13	52.0	4500	2	US-08-743-637B-35	Sequence 35, Appl
c 41	13	52.0	4500	3	US-08-526-840B-35	Sequence 35, Appl
c 42	13	52.0	5590	4	US-09-050-159-129	Sequence 129, App
c 43	13	52.0	72928	3	US-09-009-913-1	Sequence 1, Appl
c 44	12	48.0	17	1	US-08-373-124A-284	Sequence 284, App
c 45	12	48.0	17	1	US-08-373-124A-286	Sequence 286, App

ALIGNMENTS

RESULT 1
US-08-401-068-7
Sequence 7, Application US/08401068
Patent No. 5859335
GENERAL INFORMATION:
APPLICANT: Patton, David
TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/401,068
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/351,970
FILING DATE: 08-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8614
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1038
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /product= "Biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 attgtcgaagtcacagaattattt 25
 ||||||||||||||||||
 Db 24 ATTGTGCAAGTCACAGATTATT 48

RESULT 2
 US-08-846-338-7
 ; Sequence 7, Application US/08846338
 ; Patent No. 5869719
 ; GENERAL INFORMATION:
 ; APPLICANT: Patton, David
 ; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. 5869719artis Corporation
 ; STREET: 520 White Plains Road, P.O. Box 2005
 ; CITY: Tarrytown
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 10591
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30B
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/846.338
 ; FILING DATE:
 ; CLASSIFICATION: 800
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Meigs, J. Timothy
 ; REGISTRATION NUMBER: 38,241
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 919-541-8587
 ; TELEFAX: 919-541-8689
 ; INFORMATION FOR SEO ID NO: 7;
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1041 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHEICAL: NO
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1..1038
 ; IDENTIFICATION METHOD: experimental
 ; OTHER INFORMATION: /product= "biotin synthase"
 ; OTHER INFORMATION: /evidence= EXPERIMENTAL
 ; US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
 Best Local Similarity 100.0%; Pred. NO. 1.2e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 attgtcgaagtcacagaattattt 25
 ||||||||||||||||||
 Db 24 ATTGTGCAAGTCACAGATTATT 48

RESULT 3
 US-08-411-768B-1
 ; Sequence 1, Application US/08411768B
 ; Patent No. 6083712
 ; GENERAL INFORMATION:
 ; APPLICANT: Olwen Birch
 ; APPLICANT: Johann Brass
 ; APPLICANT: Martin Fuhrmann

APPLICANT: Nicholas Shaw
 TITLE OF INVENTION: Biotechnological Method
 TITLE OF INVENTION: of Producing Biotin
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
 STREET: 30 Rockefeller Plaza
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10112
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Wordperfect
 ; SOFTWARE: Version 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/411,768B
 ; FILING DATE: 31-March-95
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: CH 3124/92
 ; FILING DATE: 02-OCT-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: CH 2134/93
 ; FILING DATE: 15-JUL-1993
 ; INFORMATION FOR SEO ID NO: 1;
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 5872 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHEICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Escherichia coli
 ; STRAIN: DSM498
 ; IMMEDIATE SOURCE:
 ; CLONE: pBO30A-15/9
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 117..1157
 ; IDENTIFICATION METHOD: experimental
 ; OTHER INFORMATION: /codon_start= 117
 ; OTHER INFORMATION: /product= "biotin synthase"
 ; OTHER INFORMATION: /evidence= EXPERIMENTAL
 ; OTHER INFORMATION: /gene= "biob"
 ; OTHER INFORMATION: /number= 1
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 2295..3050
 ; OTHER INFORMATION: /codon_start= 2295
 ; OTHER INFORMATION: /function= "involved in pimeloyl-CoA synthesis"
 ; OTHER INFORMATION: /product= "protein"
 ; OTHER INFORMATION: /gene= "bioc"
 ; OTHER INFORMATION: /number= 3
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 3750..5039
 ; IDENTIFICATION METHOD: experimental
 ; OTHER INFORMATION: /codon_start= 3750
 ; OTHER INFORMATION: /EC_number= 2.6.1.62
 ; OTHER INFORMATION: /product= "DAPA synthase"
 ; OTHER INFORMATION: /evidence= EXPERIMENTAL
 ; OTHER INFORMATION: /gene= "DioA"
 ; OTHER INFORMATION: /number= 5
 ; OTHER INFORMATION: /standard_name= "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
 ; OTHER INFORMATION: /amino-transf."
 ; NAME/KEY: CDS

```
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 5098
OTHER INFORMATION: /function= "unknown, involved in biotin synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ORF1"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "promoter plac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name= "promoter plac"
FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name= "bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name= "b10a RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name= "ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name= "rho-independent"
OTHER INFORMATION: /transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "promoter plac"
OTHER INFORMATION: /evidence= EXPERIMENTAL
PUBLICATON INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

Query Match          100.0%; Score 25; DB 3; length 5872;
Best Local Similarity 100.0%; Pred. NO. 1.le-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 atgtcgcaagtcacagaattatt 25
Db 140 atgtcgcaagtcacagaattatt 164

RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass

APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect
SOFTWARE: Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBG30A15-9
FEATURE:
NAME/KEY: CDS
LOCATION: 1154..2308
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 1154
OTHER INFORMATION: /EC_number= 2.3.1.47
OTHER INFORMATION: /product= "KAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioF"
OTHER INFORMATION: /number= 2
OTHER INFORMATION: /standard_name= "8-Amino-7-oxononanoate synthase"
FEATURE:
NAME/KEY: CDS
LOCATION: 3043..3753
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3043
OTHER INFORMATION: /EC_number= 6.3.3.3
OTHER INFORMATION: /product= "DTB synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioD"
OTHER INFORMATION: /number= 4
OTHER INFORMATION: /standard_name= "Dethiobiotin synthase"
FEATURE:
NAME/KEY: RBS
LOCATION: 1141..1156
OTHER INFORMATION: /standard_name= "bioF RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3030..3045
OTHER INFORMATION: /standard_name= "biob RBS"
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
```

FILING DATE: 26-AUG-1986
 PUBLICATION DATE: 07-APR-1993
 US-08-411-768B-6

Query Match 100.0%; Score 25; DB 3; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 1,1e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 attgcgaagtcacagaatttt 25
 ||||||||||||||||||||
 Db 140 ATTGTCGAAGTCACAGAATTATT 164

RESULT 5
 US-07-944-259-2/c
 ; Sequence 2, Application US/07944259
 ; Patent No. 5885792

GENERAL INFORMATION:
 APPLICANT: Ifuku, Ohji
 APPLICANT: Haze, Shintaro
 APPLICANT: Kishimoto, Jiro
 APPLICANT: Nakahama, Kazuo
 TITLE OF INVENTION: BIOTIN OPERON
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Wegner, Cantor, Mueller & Player
 STREET: 1233 20th Street, N.W.
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20036-8218

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/944,259
 FILING DATE: 19920914
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
 NAME: Cantor, Herbert I.
 REGISTRATION NUMBER: 24,392
 REFERENCE/DOCKET NUMBER: P-450-23557
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-887-0400
 TELEFAX: 202-835-0605

TELEX: 440706
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 17 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: Escherichia coli
 US-07-944-259-2

Query Match 68.0%; Score 17; DB 2; Length 17;
 Best Local Similarity 100.0%; Pred. No. 0.38;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 gtccgaatcacaagt 20
 ||||||||||||
 Db 17 GTCCGAATCACAAGAT 1

RESULT 6
 US-09-166-203-48
 ; Sequence 48, Application US/09166203A

Patent No. 5968826
 GENERAL INFORMATION:
 APPLICANT: Bennett, C. Frank
 APPLICANT: Condon, Tom P.
 APPLICANT: Cowsett, Lex M.
 TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION
 FILE REFERENCE: ISPH-0323
 CURRENT APPLICATION NUMBER: US/09/166,203A
 CURRENT FILING DATE: 1998-10-05
 NUMBER OF SEQ ID NOS: 60
 SEQ ID NO 48
 LENGTH: 1771
 TYPE: DNA
 ORGANISM: Mus musculus
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (1193)..(1387)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (1709)..(1771)
 PUBLICATION INFORMATION:
 DATABASE ACCESSION NUMBER: U20788 Genbank
 DATABASE ENTRY DATE: 1996-04-18
 US-09-166-203-48

Query Match 60.0%; Score 15; DB 2; Length 1771;
 Best Local Similarity 100.0%; Pred. No. 3.4;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gcaagtcacagaatt 21
 |||||||||
 Db 312 gcaagtcacagaatt 326

RESULT 7
 5498529-7/c
 ; Patent No. 5498529
 ; APPLICANT: BERKA, THOMAS R.; FORNWALD, JAMES A.; GORNIAT,
 ; JOSELINA G.; ROSENBERG, MARTIN; STRICKLER, JAMES E.; TAYLOR, DEAN P.
 ; TITLE OF INVENTION: PROTEIN PROTEASE INHIBITORS FROM
 ; STREPTOMYCES
 ; NUMBER OF SEQUENCES: 9
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/162,506
 ; FILING DATE: 03-DEC-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 894,167
 ; FILING DATE: 03-JUN-1992
 ; APPLICATION NUMBER: 346,119
 ; FILING DATE: 16-FEB-1989
 ; APPLICATION NUMBER: 897,245
 ; FILING DATE: 18-AUG-1986
 ; SEQ ID NO: 7:
 ; LENGTH: 1050
 ; 5498529-7

Query Match 56.0%; Score 14; DB 6; Length 1050;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 aagtcacagaatta 22
 |||||||||
 Db 251 AAGTCACAGAATTA 238

RESULT 8
 US-08-372-652-14/c
 ; Sequence 14, Application US/08372652
 ; Patent No. 5932699
 ; GENERAL INFORMATION:
 ; APPLICANT: Moore, David

```

: APPLICANT: Seol, Mongi
: APPLICANT: Choi, Hwang-Sik
: TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
: TITLE OF INVENTION: POLYPEPTIDES AND RELATED MOLECULES AND METHODS
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESS: Fish & Richardson P.C.
: STREET: 225 Franklin Street, Suite 3100
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/372,652
: FILING DATE: 13-JAN-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Clark, Paul T.
: REGISTRATION NUMBER: 30,162
: REFERENCE/DOCKET NUMBER: 00786/246001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1677 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
: US-08-372-652-14

Query Match          56.0%; Score 14; DB 2; Length 1677;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcac 15
    |||||||||||
Db 630 TTGTGCGCAAGTCAC 617

RESULT 9
PCT-US95-16311-14/c
: Sequence 14, Application PC/TUS9516311
: GENERAL INFORMATION:
: APPLICANT: Moore, David
: APPLICANT: Seol, Mongi
: APPLICANT: Choi, Hwang-Sik
: TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
: TITLE OF INVENTION: POLYPEPTIDES AND RELATED MOLECULES AND METHODS
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESS: Fish & Richardson P.C.
: STREET: 225 Franklin Street, Suite 3100
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/16311
: FILING DATE:
: PRIOR APPLICATION DATA:

```

```

: APPLICATION NUMBER: 08/372,652
: FILING DATE: 13-JAN-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Clark, Paul T.
: REGISTRATION NUMBER: 30,162
: REFERENCE/DOCKET NUMBER: 00786/246001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1677 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
: PCT-US95-16311-14

Query Match          56.0%; Score 14; DB 5; Length 1677;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ttgtcgcaagtcac 15
    |||||||||||
Db 630 TTGTGCGCAAGTCAC 617

RESULT 10
US-08-372-652-6/c
: Sequence 6, Application US/08372652
: Patent No. 5932699
: GENERAL INFORMATION:
: APPLICANT: Moore, David
: APPLICANT: Seol, Mongi
: APPLICANT: Choi, Hwang-Sik
: TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING
: TITLE OF INVENTION: POLYPEPTIDES AND RELATED MOLECULES AND METHODS
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESS: Fish & Richardson P.C.
: STREET: 225 Franklin Street, Suite 3100
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/372,652
: FILING DATE: 13-JAN-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Clark, Paul T.
: REGISTRATION NUMBER: 30,162
: REFERENCE/DOCKET NUMBER: 00786/246001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1787 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
: US-08-372-652-6

```

Query Match 56.0%; Score 14; DB 2; Length 1787;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ttgtgcgaagtac 15
|||||
DB 740 TTGTGCGAAGTCAC 727

RESULT 11

PCT-US95-16311-6/c
; Sequence 6, Application PC/TUS9516311

; GENERAL INFORMATION:

; APPLICANT: Moore, David

; APPLICANT: Seol, Wongi

; APPLICANT: Choi, Hwang-Sik

; TITLE OF INVENTION: RETINOID X RECEPTOR-INTERACTING

; TITLE OF INVENTION: POLYPEPTIDES AND RELATED MOLECULES AND METHODS

; NUMBER OF SEQUENCES: 17

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 225 Franklin Street, Suite 3100

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/16311

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/372,652

; FILING DATE: 13-JAN-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Clark, Paul T.

; REGISTRATION NUMBER: 30,162

; REFERENCE/DOCKET NUMBER: 00786/246001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617/542-5070

; TELEFAX: 617/542-8906

; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1787 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA

; PCT-US95-16311-6

Query Match 56.0%; Score 14; DB 5; Length 1787;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ttgtgcgaagtac 15
|||||
DB 740 TTGTGCGAAGTCAC 727

RESULT 12

US-08-765-081-4

; Sequence 4, Application US/08765081

; Patent No. 5798260

; GENERAL INFORMATION:

; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.

; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
STREET: 2800 Pacific First Center, 1420 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA

ZIP: 98101-2347

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage

; COMPUTER: IBM PC compatible/Pentium

; OPERATING SYSTEM: MS-Windows 3.1

; SOFTWARE: Word for Windows-6.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/765,081

; FILING DATE: March 26, 1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/06994

; FILING DATE: June 7, 1995

; APPLICATION NUMBER: US 08/265,714

; FILING DATE: June 24, 1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Shelton, Dennis K.

; REGISTRATION NUMBER: 26,997

; REFERENCE/DOCKET NUMBER: CHOR-1-10286

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 1-206-682-8100; 1-206-224-0718 (direct)

; TELEFAX: 1-206-224-0779

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2091 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; DESCRIPTION: Corresponds to SEQ ID NO:1,

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Escherichia coli 0157: H7

; STRAIN: 86-24 NALR

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..2088

US-08-765-081-4

Query Match 56.0%; Score 14; DB 1; Length 2091;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 agtcacgaattat 23
|||||
DB 692 AGTCACGAATTAT 705

RESULT 13

US-09-098-082-4

; Sequence 4, Application US/09098082

; Patent No. 6040421

; GENERAL INFORMATION:

; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.

; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness

; STREET: 2800 Pacific First Center, 1420 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98101-2347

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage

; COMPUTER: IBM PC compatible/Pentium II

OPERATING SYSTEM: MS-Windows 95
SOFTWARE: Word for Windows-6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/098,082
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/765,081
FILING DATE: March 26, 1997
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: June 7, 1995
APPLICATION NUMBER: US 08/265,714
FILING DATE: June 24, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Sheiness, Diana K.
REGISTRATION NUMBER: 35,356
REFERENCE/DOCKET NUMBER: CHOR-1-12402
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)
TELEFAX: 1-206-224-0779
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2091 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
DESCRIPTION: Corresponds to SEQ ID NO:1,
DESCRIPTION: nucleotides 3036-5126
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli O157: H7
STRAIN: 86-24 NALR
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2088
US-09-098-082-4

Query Match 56.0%; Score 14; DB 3; Length 2091;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 agtcacagaattat 23
|||||
Db 692 AGTCACAGATTAT 705

RESULT 14
PCT-US95-06994-4
Sequence 4, Application PC/TUS9506994
GENERAL INFORMATION:
APPLICANT: Children's Hospital & Medical Center
APPLICANT: University of Washington
APPLICANT: Washington State University Research Foundation
APPLICANT: TARR, PHILIP I
APPLICANT: BIGLE, SIMA S
APPLICANT: BESSER, THOMAS E
APPLICANT: VARY JR, JAMES C
TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
STREET: SUITE 2800, 1420 FIFTH AVENUE
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: WA 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: 07-JUN-95
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/265,714
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: BRODERICK, THOMAS F
REGISTRATION NUMBER: 31,332
REFERENCE/DOCKET NUMBER: CHOR-18591
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682 8100
TELEFAX: (206) 224 0779
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2091 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
DESCRIPTION: corresponds to SEQ ID NO:1,
DESCRIPTION: nucleotides 3036-5126
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli O157: H7
STRAIN: 86-24 NALR
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2088
PCT-US95-06994-4

Query Match 56.0%; Score 14; DB 5; Length 2091;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 agtcacagaattat 23
|||||
Db 692 AGTCACAGATTAT 705

RESULT 15
US-08-496-631-1/c
Sequence 1, Application US/08496631
Patent No. 5728548
GENERAL INFORMATION:
APPLICANT: Bowman, Michael
TITLE OF INVENTION: STEROID RECEPTOR RRL
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/496,631
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15248
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224

TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2233 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 363..1778
US-08-496-631-1

Query Match 56.0%; Score 14; DB 1; Length 2233;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ttgtcgaagtcac 15
|||||
Db 1055 ttgtcgaagtcac 1042

RESULT 16
US-09-008-697A-7/c
; Sequence 7, Application US/09008697A
; Patent No. 6197504
; GENERAL INFORMATION:
; APPLICANT: Chow, King Lau
; TITLE OF INVENTION: USES OF MAB-21
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/008,697A
; FILING DATE: January 19, 1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Chan, Albert Mai-Kit
; REGISTRATION NUMBER: 36,479
; REFERENCE/DOCKET NUMBER: 50752-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2585 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-008-697A-7

Query Match 56.0%; Score 14; DB 4; Length 2585;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 caagtcacgaatt 21
|||||
Db 1058 caagtcacgaatt 1045

RESULT 17
US-08-765-081-1
; Sequence 1, Application US/08765081
; Patent No. 5798260
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44mb storage
; COMPUTER: IBM PC compatible/Pentium
; OPERATING SYSTEM: MS-Windows 3.1
; SOFTWARE: Word for Windows-6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/765,081
; FILING DATE: March 26, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K.
; REGISTRATION NUMBER: 26,997
; REFERENCE/DOCKET NUMBER: CHOR-1-10286
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-206-682-9100; 1-206-224-0718 (direct)
; TELEFAX: 1-206-224-0779
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli 0157:H7
; STRAIN: 86-24 NALR
; IMMEDIATE SOURCE:
; CLONE: PEAR
US-08-765-081-1

Query Match 56.0%; Score 14; DB 1; Length 8041;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 agtcacagaattat 23
|||||
Db 3727 agtcacagaattat 3740

RESULT 18
US-09-098-082-1
; Sequence 1, Application US/09098082
; Patent No. 6040421
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue

CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-3.5 inch, 1.44MB storage
COMPUTER: IBM PC compatible/Pentium II
OPERATING SYSTEM: MS-Windows 95
SOFTWARE: Word for Windows-6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/098,082
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/765,081
FILING DATE: March 26, 1997
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: June 7, 1995
APPLICATION NUMBER: US 08/265,714
FILING DATE: June 24, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Sheiness, Diana K.
REGISTRATION NUMBER: 35,356
REFERENCE/DOCKET NUMBER: CHOR-1-12402
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)
TELEFAX: 1-206-224-0779
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli O157:H7
STRAIN: 86-24 NALR
IMMEDIATE SOURCE:
CLONE: PEAR
US-09-098-082-1

Query Match 56.0%; Score 14; DB 3; Length 8041;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 agtcacagaattat 23
|||||
Db 3727 AGTCACAGATTAT 3740

RESULT 19
PCT-US95-06994-1
Sequence 1, Application PC/TUS9506994
GENERAL INFORMATION:
APPLICANT: Children's Hospital & Medical Center
APPLICANT: University of Washington
APPLICANT: Washington State University Research Foundation
APPLICANT: TARR, PHILIP I
APPLICANT: BILGE, SIMA S
APPLICANT: BESSER, THOMAS E
APPLICANT: VARY JR, JAMES C
TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
STREET: SUITE 2800, 1420 FIFTH AVENUE
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: WA 98101
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: 07-JUN-95
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/265,714
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: BRODERICK, THOMAS F
REGISTRATION NUMBER: 31,332
REFERENCE/DOCKET NUMBER: CHOR-18591
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682 8100
TELEFAX: (206) 224 0779
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli O157:H7
STRAIN: 86-24 NALR
IMMEDIATE SOURCE:
CLONE: PEAR
PCT-US95-06994-1

Query Match 56.0%; Score 14; DB 5; Length 8041;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 agtcacagaattat 23
|||||
Db 3727 AGTCACAGATTAT 3740

RESULT 20
US-08-875-847B-3/C
Sequence 3, Application US/08875847B
Patent No. 6255105
GENERAL INFORMATION:
APPLICANT: The Government of the United States of America as represented by the Secretary, Department of Health and Human Services; Callahan, Robert; Marchetti,
APPLICANT: Antonio; Buttilta, Riamma; Smith, Gilbert H.
TITLE OF INVENTION: Nucleotide And Deduced Amino Acid Sequences of A New Tumor Gene, Int6, And the Use Of Reagents Derived From These Sequences In Diagnostic Assays, Vaccines, Immunotherapy And Gene Therapy
TITLE OF INVENTION: Int6, And the Use Of Reagents Derived From These Sequences In Diagnostic Assays, Vaccines, Immunotherapy And Gene Therapy
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MS WORD 97
CURRENT APPLICATION DATA:

```

? APPLICATION NUMBER: US/08/875,847B
? FILING DATE: 09-FEB-1996
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/385,998
? FILING DATE: 09-FEB-1995
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: William S. Feller
? REGISTRATION NUMBER: 26,728
? REFERENCE/DOCKET NUMBER: 2026-4179PCT
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (212) 758-4800
? TELEFAX: (212) 751-6849
? TELEX: 421792
? INFORMATION FOR SEQ. ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1500 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
?
? US-08-875-847B-3

```

```

Query Match          52.0%; Score 13; DB 4; Length 1500;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 12 tcacagaattatt 24
   |||
Db 624 TCACAGATTATT 612

```

```

RESULT 21
US-08-875-847B-1/c
? Sequence 1, Application US/08875847B
? Patent No. 6255105
? GENERAL INFORMATION:
? APPLICANT: The Government of the United
? APPLICANT: States of America as represented by the
? APPLICANT: Secretary, Department of Health and Human
? APPLICANT: Services; Callahan, Robert; Marchetti,
? APPLICANT: Antonio; Buttila, Flamma; Smith, Gilbert H.
? TITLE OF INVENTION: Nucleotide And Deduced
? TITLE OF INVENTION: Amino Acid Sequences Of A New Tumor Gene,
? TITLE OF INVENTION: Int6, And The Use Of Reagents Derived From
? TITLE OF INVENTION: These Sequences In Diagnostic Assays,
? NUMBER OF SEQUENCES: 32
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
? STREET: 345 PARK AVENUE
? CITY: NEW YORK
? STATE: NEW YORK
? COUNTRY: USA
? ZIP: 10154
? COMPUTER READABLE FORM:
? MEDIUM TYPE: FLOPPY DISK
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: MS WORD 97
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/875,847B
? FILING DATE: 09-FEB-1996
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/385,998
? FILING DATE: 09-FEB-1995
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: William S. Feller
? REGISTRATION NUMBER: 26,728
? REFERENCE/DOCKET NUMBER: 2026-4179PCT

```

```

? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (212) 758-4800
? TELEFAX: (212) 751-6849
? TELEX: 421792
? INFORMATION FOR SEQ. ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1505 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
?
? US-08-875-847B-1

```

```

Query Match          52.0%; Score 13; DB 4; Length 1505;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 12 tcacagaattatt 24
   |||
Db 630 TCACAGATTATT 618

```

```

RESULT 22
US-08-651-572-1/c
? Sequence 1, Application US/08651572
? Patent No. 5789228
? GENERAL INFORMATION:
? APPLICANT: Lam, D. et al.
? TITLE OF INVENTION: Endoglucanases
? NUMBER OF SEQUENCES: 6
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: CARELLA, BYRNE, BAIN, GIUFFILAN,
? ADDRESS: CECCHI, STEWART & OLSTEIN
? STREET: 6 BECKER FARM ROAD
? CITY: ROSELAND
? STATE: NEW JERSEY
? COUNTRY: USA
? ZIP: 07068
? COMPUTER READABLE FORM:
? MEDIUM TYPE: 3.5 INCH DISKETTE
? OPERATING SYSTEM: MS-DOS
? SOFTWARE: WORD PERFECT 5.1
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/651,572
? FILING DATE: Herewith
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Charles J. Heron
? REGISTRATION NUMBER: 28,019
? REFERENCE/DOCKET NUMBER: 331400-48
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 201-994-1700
? TELEFAX: 201-994-1744
? INFORMATION FOR SEQ. ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1662 NUCLEOTIDES
? TYPE: NUCLEIC ACID
? STRANDEDNESS: SINGLE
? TOPOLOGY: LINEAR
? MOLECULE TYPE: DNA
?
? US-08-651-572-1

```

```

Query Match          52.0%; Score 13; DB 1; Length 1662;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 9 aagtcacagaatt 21
   |||

```

Db 974 AAGTCACAGAATT 962

RESULT 23

US-09-066-544-1/c
Sequence 1, Application US/09066544
Patent No. 6001984
GENERAL INFORMATION:
APPLICANT: Lam, D. et al.
TITLE OF INVENTION: Endoglucanases
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: CECCHI, STEWART & OLSTEIN,
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/066,544
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/651,572
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Charles J. Herron
REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 331400-48
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ. ID NO.: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1662 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: DNA
US-09-066-544-1

Query Match 52.0%; Score 13; DB 3; Length 1662;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 aagtcacagaatt 21
|||||

Db 974 AAGTCACAGAATT 962

RESULT 24

US-08-951-086-1/c
Sequence 1, Application US/08951086
Patent No. 6074867
GENERAL INFORMATION:
APPLICANT: Lam, D. et al.
TITLE OF INVENTION: Endoglucanases
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: CECCHI, STEWART & OLSTEIN,
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/951,086
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/651,572
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Charles J. Herron
REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 331400-48
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ. ID NO.: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1662 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: DNA
US-08-951-086-1

Query Match 52.0%; Score 13; DB 3; Length 1662;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 aagtcacagaatt 21
|||||

Db 974 AAGTCACAGAATT 962

RESULT 25

US-08-484-993B-1/c
Sequence 1, Application US/0848493B
Patent No. 5837497
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunocoreception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,993B
FILING DATE: 09-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107

REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2214 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sus scrofa
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 12..119
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 120..2153
FEATURE:
NAME/KEY: CDS
LOCATION: 12..2153
US-08-484-993B-1

Query Match 52.0%; Score 13; DB 2; Length 2214;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 7 gcaagtcacagaa 19
|||||
Db 1898 GCAAGTCACAGAA 1886

RESULT 26
US-08-484-158B-1/c
Sequence 1, Application US/08484158B
Patent No. 5976545
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Pharmaceutical Compositions for
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,158B
FILING DATE: 07-JUNE-95
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/149,223
FILING DATE: 09-NOV-93
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-93
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-92
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 32794
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2214 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sus scrofa
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 12..119
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 120..2153
FEATURE:
NAME/KEY: CDS
LOCATION: 12..2153
US-08-484-158B-1

Query Match 52.0%; Score 13; DB 2; Length 2214;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 7 gcaagtcacagaa 19
|||||
Db 1898 GCAAGTCACAGAA 1886

RESULT 27
US-08-484-596A-1/c
Sequence 1, Application US/08484596A
Patent No. 5981228
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunocontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,596A

```

: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/149,223
: FILING DATE: 11-NOV-1993
: APPLICATION NUMBER: 07/973,341
: FILING DATE: 09-NOV-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Clough, David W.
: REGISTRATION NUMBER: 36,107
: REFERENCE/DOCKET NUMBER: 31745
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312/474-6653
: TELEFAX: 312/474-0448
: TELEX: 25-3856
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2214 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Sus scrofa
: DEVELOPMENTAL STAGE: Juvenile
: HAPLOTYPE: Diploidy
: TISSUE TYPE: Ovary
: CELL TYPE: Oocyte
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 12..119
: FEATURE:
: NAME/KEY: mat_peptide
: LOCATION: 120..2153
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 12..2153
: US-08-484-596A-1

Query Match          52.0%; Score 13; DB 2; Length 2214;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      7 gcaagtcacagaa 19
      |||
Db      1898 GCAAGTCACAGAA 1886

RESULT 28
US-08-480-150A-1/c
: Sequence 1, Application US/08480150A
: Patent No. 5989550
: GENERAL INFORMATION:
: APPLICANT: Harris Ph.D., Jeffrey D.
: APPLICANT: Hsu, Kuang T.
: APPLICANT: Podolski, Joseph S.
: TITLE OF INVENTION: Materials and Methods for Immunocontraception
: NUMBER OF SEQUENCES: 59
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 6300 Sears Tower, 233 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606-6402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/480,150A
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/149,223
: FILING DATE: 09-NOV-1993
: APPLICATION NUMBER: 08/012,990
: FILING DATE: 29-JAN-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/973,341
: FILING DATE: 09-NOV-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Clough, David W.
: REGISTRATION NUMBER: 36,107
: REFERENCE/DOCKET NUMBER: 31745
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312/474-6653
: TELEFAX: 312/474-0448
: TELEX: 25-3856
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2214 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Sus scrofa
: DEVELOPMENTAL STAGE: Juvenile
: HAPLOTYPE: Diploidy
: TISSUE TYPE: Ovary
: CELL TYPE: Oocyte
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 12..119
: FEATURE:
: NAME/KEY: mat_peptide
: LOCATION: 120..2153
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 12..2153
: US-08-480-150A-1

Query Match          52.0%; Score 13; DB 2; Length 2214;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      7 gcaagtcacagaa 19
      |||
Db      1898 GCAAGTCACAGAA 1886

RESULT 29
US-08-458-731-1/c
: Sequence 1, Application US/08458731
: Patent No. 6001599
: GENERAL INFORMATION:
: APPLICANT: Harris Ph.D., Jeffrey D.
: APPLICANT: Hsu, Kuang T.
: APPLICANT: Podolski, Joseph S.
: TITLE OF INVENTION: Materials and Methods for Immunocontraception
: NUMBER OF SEQUENCES: 59
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 6300 Sears Tower, 233 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
```

ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,731
FILING DATE: 09-NOV-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEO ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2214 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sus scrofa
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 12..119
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 120..2153
FEATURE:
NAME/KEY: CDS
LOCATION: 12..2153
US-08-458-731-1

Query Match 52.0%; Score 13; DB 3; Length 2214;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 gcaagtcacagaa 19
|||||
Db 1898 GCAAGTCACAGAA 1886

RESULT 30
US-08-149-223A-1/c
Sequence 1, Application US/08149223A
Patent No. 6027727
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunoccontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/149,223A
FILING DATE: 09-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEO ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2214 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sus scrofa
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 12..119
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 120..2153
FEATURE:
NAME/KEY: CDS
LOCATION: 12..2153
US-08-149-223A-1

Query Match 52.0%; Score 13; DB 3; Length 2214;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 gcaagtcacagaa 19
|||||
Db 1898 GCAAGTCACAGAA 1886

RESULT 31
US-08-484-993B-9/c
Sequence 9, Application US/08484993B
Patent No. 5837497
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunoccontraception
NUMBER OF SEQUENCES: 59

```

CORRESPONDENCE ADDRESS:
ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,993B
FILING DATE: 09-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Canis familiaris
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 206..2353
US-08-484-993B-9

Query Match          52.0%; Score 13; DB 2; Length 2381;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 gcaagtcacagaa 19
        |||||||||||
Db      2098 GCAAGTCACAGAA 2086

RESULT 32
US-08-484-158B-9/c
; Sequence 9, Application US/08484158B
; Patent No. 5976545
; GENERAL INFORMATION:
; APPLICANT: Harris Ph.D., Jeffrey D.
; APPLICANT: Hsu, Kuang T.
; APPLICANT: Podolski, Joseph S.
; TITLE OF INVENTION: Pharmaceutical Compositions for
; TITLE OF INVENTION: Immunocounteraction
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
```

```

ADDRESS: Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,158B
FILING DATE: 07-JUNE-95
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/149,223
FILING DATE: 09-NOV-93
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-93
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-92
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 32794
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Canis familiaris
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 206..2353
US-08-484-158B-9

Query Match          52.0%; Score 13; DB 2; Length 2381;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 gcaagtcacagaa 19
        |||||||||||
Db      2098 GCAAGTCACAGAA 2086

RESULT 33
US-08-484-596A-9/c
; Sequence 9, Application US/08484596A
; Patent No. 5981228
; GENERAL INFORMATION:
; APPLICANT: Harris Ph.D., Jeffrey D.
; APPLICANT: Hsu, Kuang T.
; APPLICANT: Podolski, Joseph S.
; TITLE OF INVENTION: Materials and Methods for Immunocounteraction
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
```

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,596A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/149,223
FILING DATE: 11-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Canis familiaris
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 206..2353
US-08-484-596A-9

Query Match 52.0%; Score 13; DB 2; Length 2381;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gcaagtcacagaa 19
|||||
Db 2098 GCAAGTCACAGAA 2086

RESULT 34
US-08-480-150A-9/C
Sequence 9, Application US/08480150A
Patent No. 5989550
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunocontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago

STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,150A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/149,223
FILING DATE: 09-NOV-1993
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Canis familiaris
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 206..2353
US-08-480-150A-9

Query Match 52.0%; Score 13; DB 2; Length 2381;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gcaagtcacagaa 19
|||||
Db 2098 GCAAGTCACAGAA 2086

RESULT 35
US-08-458-731-9/C
Sequence 9, Application US/08458731
Patent No. 6001599
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunocontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois

COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,731
FILING DATE: 09-NOV-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Canis familiaris
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 206..2353
US-08-458-731-9

Query Match 52.0%; Score 13; DB 3; Length 2381;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gcaagtcacagaa 19
|||||

Db 2098 GCAAGTCACAGAA 2086

RESULT 36
US-08-149-223A-9/c
Sequence 9, Application US/08149223A
Patent No. 6027727
GENERAL INFORMATION:
APPLICANT: Harris Ph.D., Jeffrey D.
APPLICANT: Hsu, Kuang T.
APPLICANT: Podolski, Joseph S.
TITLE OF INVENTION: Materials and Methods for Immunocontraception
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/149,223A
FILING DATE: 09-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/012,990
FILING DATE: 29-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,341
FILING DATE: 09-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 31745
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6653
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2381 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Canis familiaris
DEVELOPMENTAL STAGE: Juvenile
HAPLOTYPE: Diploidy
TISSUE TYPE: Ovary
CELL TYPE: Oocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 206..2353
US-08-149-223A-9

Query Match 52.0%; Score 13; DB 3; Length 2381;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gcaagtcacagaa 19
|||||

Db 2098 GCAAGTCACAGAA 2086

RESULT 37
US-07-718-575-5/c
Sequence 5, Application US/07718575
Patent No. 5202257
GENERAL INFORMATION:
APPLICANT: Heinemann Ph.D., Stephen F.
APPLICANT: Boulter Ph.D., James R.
APPLICANT: Hollmann Ph.D., Michael MN
APPLICANT: Bettler Ph.D., Bernhard MN
APPLICANT: Jensen Ph.D., Jan E.
TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND METHODS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
STREET: 444 So. Flower St., Suite 2000
CITY: Los Angeles
STATE: California
COUNTRY: United States
ZIP: 90071-2921
COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/718,575
FILING DATE: 19910813
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Reiter Ph.D., Stephen E.
REGISTRATION NUMBER: 31192
REFERENCE/DOCKET NUMBER: P31 8962
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
TELEX: 910330318
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: Glur3
FEATURE:
NAME/KEY: CDS
LOCATION: 167..2833
US-07-718-575-5

Query Match 52.0%; Score 13; DB 1; Length 3083;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 attgtcgcaagtc 13
Db 3033 ATTGTGCGCAAGTC 3021

RESULT 38
US-08-481-206-5/c
Sequence 5, Application US/08481206
Patent No. 5739291
GENERAL INFORMATION:
APPLICANT: Heinemann Ph.D., Stephen F.
APPLICANT: Boulter Ph.D., James R.
APPLICANT: Hollmann Ph.D., Michael MNM
APPLICANT: Bettler Ph.D., Bernhard MNM
APPLICANT: Jensen Ph.D., Jan E.
TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
STREET: 444 So. Flower St., Suite 2000
CITY: Los Angeles
STATE: California
COUNTRY: United States
ZIP: 90071-2921
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/481,206
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/013,767
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Reiter Ph.D., Stephen E.
REGISTRATION NUMBER: 31192
REFERENCE/DOCKET NUMBER: P31 8962
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
TELEX: 910330318
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: Glur3
FEATURE:
NAME/KEY: CDS
LOCATION: 167..2833
US-08-481-206-5

Query Match 52.0%; Score 13; DB 1; Length 3083;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 attgtcgcaagtc 13
Db 3033 ATTGTGCGCAAGTC 3021

RESULT 39
US-08-486-269A-5/c
Sequence 5, Application US/08486269A
Patent No. 5945509
GENERAL INFORMATION:
APPLICANT: Heinemann, Stephen F.
APPLICANT: Boulter, James R.
APPLICANT: Hollmann, Michael
APPLICANT: Bettler, Bernhard
APPLICANT: Jensen, Jan E.
TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gray Cary Ware & Freidenrich LLP
STREET: 4365 Executive Drive, Suite 1600
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows DEMONSTRATION Version 2.0D
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,269A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/013,767
FILING DATE: 04-FEB-1993
APPLICATION NUMBER: 07/718,575
FILING DATE: 21-JUN-1991
APPLICATION NUMBER: PCT/US90/06153
FILING DATE: 25-OCT-1990
APPLICATION NUMBER: 07/428,116
FILING DATE: 27-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9986

TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-677-1409
TELEFAX: 619-677-1465
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
IMMEDIATE SOURCE:
CLONE: GLUR3
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 167...2830
OTHER INFORMATION:
US-08-486-269A-5

Query Match 52.0%; Score 13; DB 2; Length 3083;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 attgtcgaagtc 13
|||||
Db 3033 ATTGTCGCAAGTC 3021

RESULT 40
US-08-743-637B-35/c
Sequence 35, Application US/08743637B
Patent No. 5994066
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: PICARD, Francois J.
APPLICANT: QUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESS:
ADDRESSEE: QUARLES & BRADY
STREET: 411 EAST WISCONSIN AVENUE
CITY: MILWAUKEE
STATE: WISCONSIN
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/743,637B
FILING DATE: 04-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/526,840
FILING DATE: 11-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586,90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 4500 base pairs
TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic) 、
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
US-08-743-637B-35

Query Match 52.0%; Score 13; DB 2; Length 4500;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 aagtcacagaatt 21
|||||
Db 1911 AAGTCACAGAATT 1899

RESULT 41
US-08-526-840B-35/c
Sequence 35, Application US/08526840B
Patent No. 6001364
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: QUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: QUARLES & BRADY
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/526,840B
FILING DATE: 11-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/304,732
FILING DATE: 12-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586,90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 4500 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
US-08-526-840B-35

Query Match 52.0%; Score 13; DB 3; Length 4500;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 aagtcacagaatt 21

Db 1911 AAGTCACAGATT 1899

|||||

RESULT 42

US-09-050-159-129
; Sequence 129, Application US/09050159A
; Patent No. 6197505
; GENERAL INFORMATION:
; APPLICANT: Andersson, Leif T
; APPLICANT: Andersson, Maria K
; APPLICANT: Linstrom, Per H
; TITLE OF INVENTION: METHODS FOR ASSESSING CARDIOVASCULAR STATUS AND
; FILE REFERENCE: 1248/ID042
; CURRENT APPLICATION NUMBER: US/09/050,159A
; EARLIER FILING DATE: 1998-03-27
; EARLIER APPLICATION NUMBER: 60/042,930
; NUMBER OF SEQ ID NOS: 133
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 129
; LENGTH: 5390
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Angiotensin I converting enzyme, 5' region
US-09-050-159-129

Query Match 52.0%; Score 13; DB 4; Length 5590;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 tcacagaattatt 24
|||||
Db 294 tcacagaattatt 306

RESULT 43

US-09-009-913-1/C
; Sequence 1, Application US/09009913
; Patent No. 6087485
; GENERAL INFORMATION:
; APPLICANT: Axxs Pharmaceuticals, Inc.
; TITLE OF INVENTION: Asthma Related Genes
; NUMBER OF SEQUENCES: 339
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Ave, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/009,913
; FILING DATE: 21-JAN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: SEQ-4P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-327-3231
; TELEFAX: 650-327-3231

TELEX:

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 72928 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: Genomic DNA

US-09-009-913-1

Query Match 52.0%; Score 13; DB 3; Length 72928;
Best Local Similarity 100.0%; Pred. No. 32;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 aagtcacagaatt 21
|||||
Db 26365 AAGTCACAGATT 26353

RESULT 44

US-08-373-124A-284
; Sequence 284, Application US/08373124A
; Patent No. 5646042
; GENERAL INFORMATION:
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Draper, Kenneth
; APPLICANT: McSwiggen, James
; APPLICANT: Jarvis, Thale
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: TREATMENT OF RESTENOSIS AND
; NUMBER OF SEQUENCES: 2627
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/373,124A
; FILING DATE: January 13, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/245,466
; FILING DATE: May 18, 1994
; APPLICATION NUMBER: 08/192,943
; FILING DATE: February 7, 1994
; APPLICATION NUMBER: 07/987,132
; FILING DATE: December 7, 1992
; APPLICATION NUMBER: 07/936,422
; FILING DATE: August 26, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 209/035
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEEX: 67-3510
; INFORMATION FOR SEQ ID NO: 284:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-08-373-124A-284

Query Match 48.0%; Score 12; DB 1; Length 17;

Best Local Similarity 58.3%; Pred. No. 2e+02;

Matches 7; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 14 acagaattattt 25
|||||:|:|:|
Db 2 ACAGAAUUAUUU 13Db |||||:|:|:|
1 ACAGAAUUAUUU 12Search completed: October 9, 2001, 15:55:21
Job time: 13197 secRESULT 45
US-08-373-124A-286
; Sequence 286, Application US/08373124A
; Patent No. 5646042

; GENERAL INFORMATION:

; APPLICANT: Slinchcomb, Dan T.

; APPLICANT: Draper, Kenneth

; APPLICANT: McSwiggen, James

; APPLICANT: Jarvis, Thale

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR

; TITLE OF INVENTION: TREATMENT OF RESTENOSIS AND

; NUMBER OF SEQUENCES: 2627

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lyon & Lyon

; STREET: 633 West Fifth Street

; CITY: Suite 4700

; STATE: Los Angeles

; COUNTRY: California

; ZIP: 90071

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; MEDIUM TYPE: storage

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: IBM P.C. DOS 5.0

; SOFTWARE: Word Perfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/373,124A

; FILING DATE: January 13, 1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/245,466

; FILING DATE: May 18, 1994

; APPLICATION NUMBER: 08/192,943

; FILING DATE: February 7, 1994

; APPLICATION NUMBER: 07/987,132

; FILING DATE: December 7, 1992

; APPLICATION NUMBER: 07/936,422

; FILING DATE: August 26, 1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Walburg, Richard

; REGISTRATION NUMBER: 32,327

; REFERENCE/DOCKET NUMBER: 209/035

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (213) 489-1600

; TELEFAX: (213) 955-0440

; TELEX: 67-3510

; INFORMATION FOR SEQ ID NO: 286:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 17 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-373-124A-286

Query Match 48.0%; Score 12; DB 1; Length 17;

Best Local Similarity 58.3%; Pred. No. 2e+02;

Matches 7; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 14 acagaattattt 25

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:15:21 ; Search time 5323.87 seconds
(without alignments)
44.389 Million cell updates/sec

Title: US-09-396-196f-2

Perfect score: 25

Sequence: 1 attgtcgaagtcacagaattatt 25

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 9

Total number of hits satisfying chosen parameters: 563573

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*

190: gb_est110:*
191: gb_est111:*
192: gb_hic:*
193: em_gss_fung:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_invl:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrl1:*
216: em_gss_vrl2:*
217: em_gss_vrl3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inva:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	72.0	357	BE800706	BE800706 sq97c06.y
2	18	72.0	388	AW781150	AW781150 s189g06.y
3	18	72.0	751	BE658717	BE658717 GM700000A
4	17	68.0	424	AQ227679	AQ227679 HS_2012-B
5	17	68.0	944	A1529608	A1529608 u180b10.y
6	16	64.0	263	BE141316	BE141316 BBI141316
7	16	64.0	270	BE181705	BE181705 QV1-HT063
8	16	64.0	353	BA450863	BA450863 BBA50863
9	16	64.0	405	AA648449	AA648449 ns22f10.s
10	16	64.0	425	BF726110	BF726110 by01c01.x
11	16	64.0	425	BF732119	BF732119 dd9d06.x
12	16	64.0	443	A1149091	A1149091 q85f02.x
13	16	64.0	454	AQ086771	AQ086771 HS_2175.B
14	16	64.0	457	A1361386	A1361386 qy47f01.x
15	16	64.0	471	BE630303	BE630303 dab93h11.x
16	16	64.0	471	A1635259	A1635259 cz79n07.x
17	16	64.0	503	A1634722	A1634722 tz09e04.x
18	16	64.0	512	BE347513	BE347513 dab87h12.x
19	16	64.0	512	AW418528	AW418528 CITR1-E1
20	16	64.0	527	AQ472309	AQ472309 CITR1-E1
21	16	64.0	649	BF322332	BF322332 maa34h05
22	16	64.0	652	AQ541562	AQ541562 RPT1-11-3
23	16	64.0	657	BF726111	BF726111 by01c01.y
24	16	64.0	702	A2382380	A2382380 1M0139H04
25	16	64.0	753	BE496183	BE496183 602538125
26	16	64.0	971	AC090023	AC090023 HS_3180.A
27	15	60.0	214	AW871559	AW871559 r47d04.y
28	15	60.0	233	BE763893	BE763893 RC4-NT005
29	15	60.0	253	BE763892	BE763892 RC4-NT005
30	15	60.0	256	AA317110	AA317110 EST19001
31	15	60.0	261	245844	245844 HSC2VE041.n
32	15	60.0	263	BE745516	BE745516 CM2-BT082
33	15	60.0	266	AA115083	AA115083 z106c10.r
34	15	60.0	282	BF800965	BF800965 OVA-CT1010
35	15	60.0	286	AA344205	AA344205 EST50086
36	15	60.0	292	B53227	B53227 CIT-HSP-200
37	15	60.0	317	BE743951	BE743951 CM2-BT082
38	15	60.0	327	T30901	T30901 EST24536.Hu
39	15	60.0	335	BF743964	BF743964 CM2-BT082
40	15	60.0	346	BF743954	BF743954 CM2-BT082
41	15	60.0	370	T07051	T07051 EST01940.Fe
42	15	60.0	381	HS3286	HS3286 Y985T09.r1
43	15	60.0	395	AA229851	AA229851 nc51e03.x
44	15	60.0	402	A1694429	A1694429 wd83f06.x
45	15	60.0	404	A1203847	A1203847 qf75h03.x

ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
1	BE800706	sq97c06.y1 Gm-c1049 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1049-1091 5' similar to TR:082150 082150 CHLOROPLAST FTSH	BE800706	357 bp	mrna	EST	20-SEP-2000		
	BE800706	PROTEASE, , mRNA sequence.	BE800706	GI:10231818					
	BE800706	EST.							
	soybean.								
	Glycine max								
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;								
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;								
	Rosidae; eustoids I; Fabales; Fabaceae; Papilionoideae; Glycine.								
	1 (bases 1 to 357)								
	Shoemaker,R., Keim,P., Vodkin,L., Eipelting,J., Coryell,V., Khanna								
	,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,								

JOURNAL
TITLE
COMMENT

Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
,T., Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk
,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R./Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Putative full length read
vector to vector length is 364
high quality sequence stop: 340.
Location/Qualifiers

FEATURES

source

1..357
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1049-1091"
/clone_lib="Gm-c1049"
/tissue_type="whole seedlings of greenhouse grown plants"
/dev_stage="3 week old"
/lab_host="DH10B"
/note="Vector: pBluescript II SK+, Site_1: EcoRI; Site_2:
XhoI; The Clark NIL was constructed and seed was provided
by Dr. J. Specht, University of Nebraska (Shoemaker and
Specht, 1995). The cDNA library was constructed from mRNA
isolated from whole seedlings of 3 week old greenhouse
grown plants. Complementary DNA was synthesized from mRNA
using a primer consisting of a poly(dT) sequence with a
XhoI restriction site and a 3' anchor. EcoRI adapters were
ligated to the blunt-ended cDNA fragments followed by XhoI
digestion. The cDNA fragments were directionally cloned
into the EcoRI-XhoI restriction site of the pBluescript
vector. The ligated cDNA fragments were transformed into
DH10B host cells (GibcoBRL). The library was constructed
in cooperation with Dr. Paul Keim's laboratory at Northern
Arizona University."

BASE COUNT

169 a 53 c 47 g 88 t

Query Match 72.0%; Score 18; DB 140; Length 357;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ccgaagtcacagaatcat 23
|||||
Db 138 CGCAAGTCACAGAAATAT 155

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE
2	AW781150	s189g06.y1 Gm-c1037 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1037-1235 5' similar to TR:082150 082150 CHLOROPLAST FTSH	AW781150	388 bp	mrna	EST	12-MAY-2000	
	AW781150	PROTEASE, , mRNA sequence.	AW781150	GI:7795753				
	AW781150	EST.						
	soybean.							
	Glycine max							
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;							
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;							
	Rosidae; eustoids I; Fabales; Fabaceae; Papilionoideae; Glycine.							
	1 (bases 1 to 388)							

AUTHORS

Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corryell, V., Khanna, A., Balla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Powers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

TITLE

Public Soybean EST Project

JOURNAL

Unpublished (1999)

COMMENT

Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
High quality sequence stop: 375.

FEATURES

SOURCE

Location/Qualifiers
1. 388
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl037-1235"
/clone_1id="Gm-cl037"
/tissue_type="fully expanded leaves of greenhouse grown plants"
/dev_stage="2 week old"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This cDNA library was constructed from mRNA isolated from fully expanded leaves of greenhouse grown plants that were 2 weeks old. The library was prepared using the Life Technologies pSPORT1 cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly(dI) sequence with a NotI restriction site. SalI linker adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli Electro-Max DH10B host cells. This library was constructed in the laboratory of Dr. Lila Vodkin by Anu Khanna at the University of Illinois at Urbana-Champaign. email: l-vodkin@uiuc.edu"

BASE COUNT

123 a 67 c 92 g 106 t

ORIGIN

Query Match 72.0%; Score 18; DB 120; Length 388;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 cgcgaatcacagaattat 23
|||||

Db 334 CGCAGTCCACAGATTAT 351

RESULT

3

BE658717/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

BE658717 751 bp mRNA EST 06-SEP-2000
GM700007A10F9 Gm-r1070 Glycine max cDNA clone Gm-r1070-2561 3',
mRNA sequence.
BE658717
BE658717.1 GI:9984609
EST.
soybean.
Glycine max
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:
Rosidae: eurosids 1: Fabales: Fabaceae: Papilionoideae: Glycine.
1 (bases 1 to 751)

AUTHORS

Vodkin, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Corryell, V., Erpelting, J., Raph, C., Shoop, E., Pardinas, J., Liu, L. and Lewin, H.
A Functional Genomics Program for Soybean (NSF 9872565)
Unpublished (1999)

JOURNAL

Unpublished (1999)

COMMENT

Other ESTs: AM156684 corresponding to Gm-cl015-2632 (5')
Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.

FEATURES

SOURCE

Location/Qualifiers
1. 751
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="Gm-r1070-2561"
/clone_1id="Gm-r1070"
/note="The library Gm-r1070 is a sequence-driven, rerecked set of 9,216 clones selected from cDNA libraries from various tissues and stages of development of soybean that represent 2,639 sequences from immature cotyledons, 1,770 from immature seed coats, 3,938 from flowers, and 869 from young pods. The 5' ESTs of the source clones from the different libraries was used to select singletons, or a representative of each contig, which were rerecked to form library Gm-r1070. The cDNA clones of the rerecked Gm-r1070 library were then sequenced at the 3' end. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, http://www.bcb.umn.edu/ResearchProjects/soybean/index.html
Rereckling was performed by genome systems, St. Louis, http://www.genomesystems.com, and 3' sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, http://www.life.uiuc.edu/ditech/keck.html. Note: The corresponding 5' EST from each clone in the Gm-r1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the genome systems clone ID of the original cDNA library that is also listed under 'OTHER EST'."

BASE COUNT

193 a 171 c 131 g 226 t 30 others

ORIGIN

Query Match 72.0%; Score 18; DB 138; Length 751;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 cgcgaatcacagaattat 23
|||||

Db 112 CGCAGTCCACAGATTAT 95

RESULT

4

AQ227679

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

human.

AQ227679 424 bp DNA GSS 26-SEP-1998
HS_2012_B2_F10_T7 CIT Approved Human Genomic Sperm Library D Homo
sapientis genomic clone plate=2012 Col=20 Row=L, DNA sequence.
AQ227679
AQ227679.1 GI:3652908
GSS.

MGI:973023	
Seq primer: custom primer used	
High quality sequence stop: 502.	
FEATURES	Location/Qualifiers
source	1. .944
	/organism="Mus musculus"
	/strain="C57BL"
	/db_xref="taxon:10090"
	/clone="IMAGE:1888699"
	/clone_11b="Sugano mouse liver ml1a"
	/sex="female"
	/dev_stage="adult"
	/lab_host="DH10v"
	/note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dt) primer [ATGCGCCCTTTTCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor (TGTGGCCCTGCTG) and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTCTGCGCTTAAAGCTCG and 3' end primer CGACCTGCACCTGAGACAC."
BASE COUNT	253 a 238 c 226 g 227 t
ORIGIN	
Query Match	68.0%; Score 17; DB 21; Length 944;
Best Local Similarity	100.0%; Prod. No. 7.1;
Matches 17; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY 9	aagtcacagaattattt 25
Db 906	AAGTCACAGAATTATT 890
RESULT 6	
BM141316/c	
LOCUS	BM141316 263 bp mRNA EST 28-JUN-2000
DEFINITION	BM141316 RIKEN full-length enriched, adult female vagina Mus musculus cDNA clone 993007004 3', mRNA sequence.
ACCESSION	BM141316
VERSION	BM141316.1 GI:8796253
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 263)
AUTHORS	Kono, H., Atawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carrinci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogaue, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomiina, N., Toya, T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
JOURNAL	RIKEN Mouse ESTs (Kono, H., et al.)
COMMENT	Unpublished (2000) Contact: Yoshihide Hayashizaki Genome Exploration Research Group, Life Science Tsukuba Center, Genome Science Laboratory The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan Tel: +81-298-36-9013

TITLE
JOURNAL
COMMENT

M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Konno, H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyada, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp,
URL: http://genome.rtc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki,
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermolabile and thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Katsunari, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.

FEATURES
source

Location/Qualifiers
1..299
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D13006117"
/clone_lib="RIKEN full-length enriched, 12 days embryo
spinal ganglion"
/tissue_type="spinal ganglion"
/dev_stage="12 days embryo"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI. cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGAGAGAGCGCGCCGACACTCGACTGTTTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5',
GAGAGAGAGATTCGACTTAAATTATATCCCGCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT
ORIGIN
75 a 72 c 45 g 107 t

Query Match
Best Local Similarity 64.0%; Score 16; DB 134; Length 299;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

LOCUS
DEFINITION
ACCESSION
AA648449
353 bp mRNA
EST
13-NOV-1997
ns22f10.s1 NCI-CGAP, GCB1 Homo sapiens cDNA clone IMAGE:1184395 3'
similar to contains element LTR5 repetitive element ;, mRNA
sequence.
AA648449

VERSION
KEYWORDS
SOURCE
ORGANISM

AA648449.1 GI:2574878
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 353)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Insert Length: 2027 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 334.

FEATURES
source

Location/Qualifiers
1..353
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1184395"
/clone_lib="NCI-CGAP, GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allan
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTACCAATCTCAAGTGGAGCGCGCCGCTCAATTTTCTTTTCTTTT-3',
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT
ORIGIN
131 a 49 c 39 g 134 t

Query Match
Best Local Similarity 64.0%; Score 16; DB 10; Length 353;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BF726110
405 bp mRNA
EST
05-JAN-2001
by01c01.x1 Human lens cDNA (Un-normalized, unamplified); BY Homo
sapiens cDNA clone by01c01 3', mRNA sequence.

REFERENCE
1 (bases 1 to 405)
Wistow, G.J., Bernstein, S., Behal, A. and Smith, D.
NEIBANK: EST analysis and bioinformatics for ocular genomics
Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
Contact: Wistow, G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078

LOCUS
DEFINITION
ACCESSION
AA648449
353 bp mRNA
EST
13-NOV-1997
ns22f10.s1 NCI-CGAP, GCB1 Homo sapiens cDNA clone IMAGE:1184395 3'
similar to contains element LTR5 repetitive element ;, mRNA
sequence.
AA648449

FEATURES
source
Location/Qualifiers
1. 405
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Hy01c01"
/clone_lib="Human Lens cDNA (Un-normalized, unamplified):
BY"
/tissue_type="Lens"
/dev_stage="Adult"
/lab_host="EMD10B"
/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
from different adults (both approximately 40 years old)
together yielded 20ug of total RNA and 150ng mRNA for cDNA
library synthesis. A directionally cloned cDNA library in
the pCMVSPORT6 vector was constructed at Life Technologies
essentially following the protocols of the Superscript
Plasmid System full details of which are contained in the
manufacturer's instruction manual
(http://www.lifetech.com/). First strand synthesis was
carried out using a Not I primer-adaptor
[5'-PGACTAGTCTAGATCGGAGCGCCGCC(TT)15-3']. Not I/blunt
end inserts were cloned into the Not I/EcoR V sites in the
vector. EST analysis was performed on the unamplified
library at the NIH Intramural Sequencing Center (NISC)."

BASE COUNT 140 a 58 c 47 g 160 t
ORIGIN

Query Match 64.0%; Score 16; DB 168; Length 405;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 caagtcacagaattat 23
|||||
Db 249 CAAGTCACAGAAATTAT 264

RESULT 11
LOCUS BF232119 429 bp mRNA EST 11-DEC-2000
DEFINITION dd99dd06.x1 Wellcome CRC PRN3 oocyte Xenopus laevis cDNA clone
IMAGE:3436475 3', mRNA sequence.
ACCESSION BF232119
VERSION BF232119.1 GI:11174440
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 429)
Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D.,
Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person,
B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,
Waterston, R. and Wilson, R.
Washu Xenopus EST project, 1999
Unpublished (1999)
Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Library constructed by N. Garrett, P. Lemaire, A.M. Zorn, and J.B.
Gurdon (Wellcome/CRC Institute). DNA Sequencing by: Washington
University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.C.E. Consortium/LLNL at: info@image.llnl.gov
High quality sequence stop: 425.

FEATURES
source
Location/Qualifiers
1. 429
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:3436475"
/clone_lib="Wellcome CRC PRN3 oocyte"
/tissue_type="oocyte"
/lab_host="DH10B (phage-resistant)"
/note="Vector: pBSRN3; Site 1: NotI; Site 2: EcoRI; cDNAs
were oligo-dT primed and directionally cloned. Staging
according to Neuwoop and Faber. Library was constructed
by N. Garrett, P. Lemaire, A.M. Zorn, and J.B. Gurdon
(Wellcome/CRC Institute)."

BASE COUNT 141 a 73 c 74 g 141 t
ORIGIN

Query Match 64.0%; Score 16; DB 146; Length 429;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 agtcacagaattatt 25
|||||
Db 60 AGTCACAGAAATTATT 75

RESULT 12
LOCUS A1149091 443 bp mRNA EST 28-OCT-1998
DEFINITION gc85f02.x1 Soares.pregnant_uterus_NbHPU Homo sapiens cDNA clone
IMAGE:1721019 3', mRNA sequence.
ACCESSION A1149091
VERSION A1149091.1 GI:3677560
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 443)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps@email.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 527 Std Error: 0.00
Seq primer: -40m13 fwd. ET from AmerSham.
Location/Qualifiers
1. 443
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1721019"
/clone_lib="Soares.pregnant_uterus_NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT73-Pac; Site 1: Not I;
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5'
AAGTCGAGCAATTCGCGCGCCGCTTTTCTTTTCTTTTCTTTT 3']
(pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

FEATURES
source
Location/Qualifiers
1. 443
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1721019"
/clone_lib="Soares.pregnant_uterus_NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT73-Pac; Site 1: Not I;
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5'
AAGTCGAGCAATTCGCGCGCCGCTTTTCTTTTCTTTTCTTTT 3']
(pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

BASE COUNT 126 a 122 c 71 g 124 t
ORIGIN

Query Match 64.0%; Score 16; DB 16; Length 443;
Best Local Similarity 100.0%; Pred. No. 26;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 agtcacagaattat 25
|||||

Db 100 AGTCACAGAATTATTT 115

RESULT 13
LOCUS A0086771 454 bp DNA GSS 26-AUG-1998
DEFINITION HS_2175_B2_A06_MR_CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone plate=2175 Col=12 Row=B, DNA sequence.
ACCESSION A0086771
VERSION A0086771.1 GI:3455988
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 454)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,U., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2175 Row: B Column: 12
Class: BAC ends
High quality sequence stop: 454.
Location/Qualifiers
1.454
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2175 Col=12 Row=B"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: Sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 188 a 70 c 76 g 117 t 3 others
ORIGIN

Query Match 64.0%; Score 16; DB 224; Length 454;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 caagtcacagaattat 23
|||||

Db 156 CAAGTCACAGAATTAT 171

RESULT 14
LOCUS A1361386 457 bp mRNA EST 15-FEB-1999
DEFINITION QY47F01.X1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2015161 3',
mRNA sequence.
ACCESSION A1361386
VERSION A1361386.1 GI:4113007
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 457)

AUTHORS NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BRGAP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-ri@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldi, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILM: at:
www.bio.lnlnl.gov/db/brp/image/image.html
Insert length: 1228 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 434.
Location/Qualifiers
1.457
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2015161"
/clone_lib="NCI CGAP_Brn23"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCGTGAAGTGGAGCGGCGCCATATCTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldi."

BASE COUNT 153 a 73 c 52 g 179 t
ORIGIN

Query Match 64.0%; Score 16; DB 19; Length 457;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 caagtcacagaattat 23
|||||

Db 153 CAAGTCACAGAATTAT 168

RESULT 15
LOCUS BG360303 467 bp mRNA EST 06-MAR-2001
DEFINITION dab93h11.y1 NICHD XCC Emb2 Xenopus laevis cDNA clone IMAGE:4405125
5', mRNA sequence.
ACCESSION BG360303
VERSION BG360303.1 GI:13240994
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 467)
AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterston,R. and Wilson,R.
Washu Xenopus EST project, 1999
Unpublished (1999)
TITLE Contact: Sandy Clifton, Ph.D.
JOURNAL Washu Xenopus EST project, 1999
COMMENT Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewatson.wustl.edu
 Library constructed by Life Technologies. DNA Sequencing by:
 Washington University Genome Sequencing Center
 Clone distribution: Xenopus clones from this library are available
 through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
 Seq primer: -40RP from Gibco
 High quality sequence stop: 465.
 Location/Qualifiers
 1..467

FEATURES
 source
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone="IMAGE:4405125"
 /clone_1ib="NICHD XGC Emb2"
 /tissue_type="embryo, stage 17/19"
 /lab_host="DH10B (phage-resistant)"
 /note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI;
 Cloned unidirectionally. Primer: Oligo dt. Average insert
 size 2.1 kb. Constructed by Life Technologies."
 BASE COUNT 146 a 86 c 84 g 151 t
 ORIGIN

Query Match 64.0%; Score 16; DB 152; Length 467;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 10 agtcacagaattatc 25
 ||||||||||||||||
 Db 89 agtcacagaattatc 104

RESULT 16
 A1635259 471 bp mRNA EST 16-DEC-1999
 LOCUS tz79h07.x1 NCI_CGAP_Pan1 Homo sapiens cDNA IMAGE:2294845 3',
 DEFINITION mRNA sequence.
 ACCESSION A1635259 GI:4686589
 VERSION A1635259
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 471)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Life Technologies catalog #: 11548-013
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 1533 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 399.
 Location/Qualifiers
 1..471

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2294845"
 /clone_1ib="NCI_CGAP_Pan1"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.72 kb. Life Technologies catalog #:
 11548-013"
 BASE COUNT 171 a 69 c 53 g 178 t

ORIGIN

Query Match 64.0%; Score 16; DB 22; Length 471;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 8 caagtcacagaattat 23
 ||||||||||||||||
 Db 235 CAAGTCACAGAAATTAT 250

RESULT 17
 A1634722 503 bp mRNA EST 16-DEC-1999
 LOCUS tz09e04.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2288094 3',
 DEFINITION similar to contains element XTR repetitive element ;, mRNA
 sequence.
 ACCESSION A1634722 GI:4686052
 VERSION A1634722
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 503)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 3719 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 413.
 Location/Qualifiers
 1..503

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2288094"
 /clone_1ib="NCI_CGAP_Ut1"
 /tissue_type="well-differentiated endometrial
 adenocarcinoma, 7 pooled tumors"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.75 kb. Life Technologies catalog #:
 11538-014"

BASE COUNT 184 a 77 c 56 g 186 t
 ORIGIN

Query Match 64.0%; Score 16; DB 22; Length 503;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 caagtcacagaattat 23
 ||||||||||||||||
 Db 236 CAAGTCACAGAAATTAT 251

RESULT 18
 BG347513 503 bp mRNA EST 28-FEB-2001
 LOCUS BG347513
 DEFINITION dab87h12.x1 NICHD XGC Emb2 Xenopus laevis cDNA clone IMAGE:4404719

3', mRNA sequence.

ACCESSION BG347513
 VERSION BG347513.1 GI:13167937
 KEYWORDS EST.
 SOURCE African clawed frog.
 ORGANISM Xenopus laevis

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae; Xenopodinae; Xenopus.

AUTHORS 1 (bases 1 to 503)
 Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Page, D., Martin, J., Wyllie, T., Underwood, K., Theising, B., Bowers, Y., Person, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R., Waterston, R., and Wilson, R.,
 Washu Xenopus EST project, 1999

TITLE Washu Xenopus EST project, 1999
 JOURNAL Unpublished (1999)
 COMMENT Contact: Sandy Clifton, Ph.D.
 Washu Xenopus EST project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Library constructed by Life Technologies. DNA sequencing by:
 Washington University Genome Sequencing Center
 Clone distribution: Xenopus clones from this library are available through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
 Seq primer: -400p from Gibco
 High quality sequence stop: 432.

FEATURES
 source
 1..503
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone="IMAGE:4404719"
 /clone_1ib="NICHD XGC Emb2"
 /tissue_type="embryo, stage 17/19"
 /lab_host="DH10B (phage-resistant)"
 /note="Vector: PCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.1 kb. Constructed by Life Technologies."

BASE COUNT 167 a 87 c 89 g 160 t

ORIGIN

Query Match 64.0%; Score 16; DB 152; Length 503;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 agtcacagaattatt 25
 ||||||||||||||||
 Db 51 agtcacagaattatt 66

RESULT 19
 AM418528 512 bp mRNA EST 09-FEB-2000
 LOCUS x132c07.x1 NCI-CGAP_U74 Homo sapiens cDNA clone IMAGE:2761836 3'
 DEFINITION Similar to contains element MER32 repetitive element ;, mRNA sequence.
 ACCESSION AM418528
 VERSION AM418528.1 GI:6946410
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 512)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
 Seq primer: -400p from Gibco
 High quality sequence stop: 404.

FEATURES
 source
 1..512
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2761836"
 /clone_1ib="NCI-CGAP_U74"
 /tissue_type="serous papillary carcinoma, high grade, 2 pooled tumors"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: PCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.48 kb. Life Technologies catalog #: 11542-016"

BASE COUNT 186 a 79 c 56 g 191 t

ORIGIN

Query Match 64.0%; Score 16; DB 115; Length 512;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 caagtcacagaattat 23
 ||||||||||||||||
 Db 240 caagtcacagaattat 255

RESULT 20
 AO472309 527 bp DNA GSS 23-APR-1999
 LOCUS CITBI-EI-2587F18.TF CITBI-EI Homo sapiens genomic clone 2587F18,
 DEFINITION DNA sequence.
 ACCESSION AO472309
 VERSION AO472309.1 GI:4655963
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 527)
 Zhao, S., Adams, M.D., Nierman, W., Malek, J., Shizuya, H., Simon, M. and Venter, J.C.,
 Use of BAC End Sequences from Caltech Libraries for Sequence-Ready Map Building
 Unpublished (1997)
 Other-GSS: CITBI-EI-2587F18.TF
 Contact: Shaying Zhao, William Nierman, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbeetlgr.org
 Clones are available from Research Genetics (info@resgen.com). BAC end search page:
 http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html.
 Seq primer: M13-21
 Class: BAC ends.

FEATURES
 source
 1..527
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="2587F18"
 /clone_1ib="CITBI-EI"

/sex="male"
/cell_type="sperm"
/note="Vector: pBelosAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC Library D"

BASE COUNT 150 a 88 c 83 g 205 t 1 others

ORIGIN

Query Match 64.0%; Score 16; DB 229; Length 527;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 agtcacagaattatt 25
|||||
Db 81 AGTCACAGAAATTATT 96

RESULT 21
BF322232 649 bp mRNA EST 21-NOV-2000
LOCUS BFA32232/c
DEFINITION maa34h05.y1 NCI-CGAP_L110 Mus musculus cDNA clone IMAGE:3812985 5' similar to SW:CPFL_MOUSE Q05421 CYTOCHROME P450 2b1 /, mRNA sequence.

ACCESSION BF322232 GI:11271563
VERSION BF322232
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 649)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaaps-remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

MG1:1455097
Seq primer: -40RP from Gibco
High quality sequence stop: 434.
FEATURES
Location/Qualifiers

1..649
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3812985"
/clone_lib="NCI-CGAP_L110"
/sex="female"
/dev_stage="10 weeks"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.6 kb. Library constructed by Life Technologies"

BASE COUNT 201 a 150 c 150 g 147 t 1 others

ORIGIN

Query Match 64.0%; Score 16; DB 147; Length 649;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 agtcacagaattatt 25
|||||
Db 517 AGTCACAGAAATTATT 502

RESULT 22
AO541562 652 bp DNA GSS 19-MAY-1999
LOCUS AO541562/c
DEFINITION RPCI-11-361M22.TV RPCI-11 Homo sapiens genomic clone RPCI-11-361M22 /, DNA sequence.

ACCESSION AO541562
VERSION AO541562.1 GI:4872018
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 652)
AUTHORS Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter, J.C.

TITLE Use of BAC end sequences from library RPCI-11 for Sequence-Ready Map Building

JOURNAL Unpublished (1997)
COMMENT Other-GSSs: RPCI-11-361M22.TV
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208

Email: hbeetlgr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs (<http://info@resgen.com>). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.

FEATURES
Location/Qualifiers

1..652
/organism="Homo sapiens"
/db_xref="GDB:7638549"
/db_xref="taxon:9606"
/clone="RPCI-11-361M22"
/clone_lib="RPCI-11"
/sex="Male"

/cell_type="Lymphocytes"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC Library"

BASE COUNT 201 a 162 c 82 g 207 t

ORIGIN

Query Match 64.0%; Score 16; DB 230; Length 652;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 caagtcacagaattat 23
|||||
Db 96 CAAGTCACAGAAATTAT 81

RESULT 23
BF726111/c 657 bp mRNA EST 05-JAN-2001
LOCUS BF726111
DEFINITION by01c01.y1 Human Lens cDNA (un-normalized, unamplified): BY Homo sapiens cDNA clone by01c01 5', mRNA sequence.

ACCESSION BF726111
VERSION BF726111.1 GI:12042022
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 657)
AUTHORS Wistow, G.J., Bernstein, S., Behal, A. and Smith, D.

TITLE NEIBANK: EST analysis and bioinformatics for ocular genomics
JOURNAL Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
COMMENT Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 01 row: c column: 01
Seq primer: M3RPI reverse primer (ABI).
Location/Qualifiers

FEATURES
source
1..657
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="by01c01"
/clone_lib="Human Lens cDNA (un-normalized, unamplified):
BY"
/tissue_type="Lens"
/dev_stage="Adult"
/lab_host="EMDH10B"
/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
from different adults (both approximately 40 years old)
together yielded 20ug of total RNA and 150ng mRNA for cDNA
library synthesis. A directionally cloned cDNA library in
the pCMVSPORT6 vector was constructed at Life Technologies
, essentially following the protocols of the Superscript
Plasmid System full details of which are contained in the
manufacturer's instruction manual
(http://www.lifetech.com/). First strand synthesis was
carried out using a Not I primer-adaptor
[5'-pGACTGATCTGATGATCGGCGCGCCG(T)15-3']. Not I/blunt
vector. EST analysis was performed on the unamplified
library at the NIH Intramural Sequencing Center (NISC)."

BASE COUNT 232 a 86 c 114 g 225 t
ORIGIN

Query Match 64.0%; Score 16; DB 168; Length 657;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 caagtcacagaattat 23
|||||
Db 574 CAAGTCACAGATTAT 559

RESULT 24
LOCUS AZ382380 702 bp DNA GSS 02-OCT-2000
DEFINITION 1M0139H04R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0139H04 R. DNA sequence.
ACCESSION AZ382380
VERSION AZ382380.1 GI:10496080
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 702)
REFERENCE
AUTHORS Islam, H., Longacre, S., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
JOURNAL Contact: Robert B. Weiss
COMMENT University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
84112, USA

Tel: 801 585 5606
Fax: 801 585 7177
Email: dduenne@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0139 row: H column: 04
Seq primer: CACACAGCAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 702.
Location/Qualifiers

FEATURES
source
1..702
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0139H04"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (91473211419b1AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 272 a 120 c 104 g 206 t
ORIGIN

Query Match 64.0%; Score 16; DB 242; Length 702;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 aagtcacagaattat 24
|||||
Db 637 AAGTCACAGATTATT 652

RESULT 25
LOCUS BG496183 753 bp mRNA EST 27-MAR-2001
DEFINITION 602538125F1 NIH-MGC_59 Homo sapiens cDNA clone IMAGE:4658982 5',
mRNA sequence.
ACCESSION BG496183
VERSION BG496183.1 GI:13457699
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 753)
REFERENCE
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@phs-remail.nih.gov
Tissue Procurement: ARCC
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
 Plate: L1CML454 row: b column: 07
 High quality sequence start: 20
 High quality sequence stop: 694.
 Location/Qualifiers
 1..753

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:4658982"
 /clone_lib="NIH_MGC_59"
 /issue_type="mucoepidermoid carcinoma"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: Lung; Vector: pDNR-LIB (Clontech); Site_1: S11 (ggcgcgcgcgcgc); Site_2: S11 (ggcgcgcgcgc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCGCATATGCCC-3' and 3' adaptor sequence: 5'-ATGCTAGAGCGCGCGCGCATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."
 BASE COUNT 202 a 143 c 174 g 233 t 1 others
 ORIGIN

Query Match 64.0%; Score 16; DB 154; Length 753;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 aagtcacagaattatt 24
 ||||||||||||||||
 Db 46 aagtcacagaattatt 31

RESULT 26
 AQ900023/C
 LOCUS
 DEFINITION HS_3180_A2_G11_T7C CIT Approved Human Genomic Sperm Library D Homo
 sapiens genomic clone Plate=3180 Col=22 Row=M, DNA sequence.
 ACCESSION AQ900023
 VERSION AQ900023.1 GI:6356213
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 911)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE
 JOURNAL Sequence-tagged connectors: A sequence approach to mapping and
 MEDLINE scanning the human genome
 COMMENT Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 99380589
 Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 3180 row: M column: 22
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 971.
 Location/Qualifiers
 1..971
 /organism="Homo sapiens"

FEATURES
 source

/db_xref="taxon:9606"
 /clone_plate="3180 Col=22 Row=M"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-coli DH10B"
 BASE COUNT 249 a 130 c 128 g 464 t
 ORIGIN

Query Match 64.0%; Score 16; DB 235; Length 971;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 aagtcacagaattatt 25
 ||||||||||||||||
 Db 501 aagtcacagaattatt 486

RESULT 27
 AM871559 214 bp mRNA EST 05-APR-2001
 LOCUS r79604.y1 Bird-Rao Meloidogyne incognita J2 Meloidogyne incognita
 DEFINITION cDNA 5' similar to WP:C26D10.4 CE03027 ;, mRNA sequence.
 ACCESSION AM871559
 VERSION AM871559.1 GI:8005612
 KEYWORDS EST.
 SOURCE southern root-knot nematode.
 ORGANISM Meloidogyne incognita
 Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
 Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

REFERENCE
 AUTHORS 1 (bases 1 to 214)
 McCarter,J., Clifton,S., Chapell,B., Pape,D., Martin,J., Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y., Gibbons,M., Ritey,E., Bennett,J., Franklin,C., Tsagarisvill,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
 The Washington Univ. Nematode EST Project, 1999
 Unpublished (1999)
 Contact: McCarter JP
 The Washington Univ. Nematode EST Project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 The library was constructed by Uma Rao and David Bird (david.bird@wustl.edu) at North Carolina State University. DNA sequencing by: Washington University Genome Sequencing Center St. Louis.

TITLE
 JOURNAL Trace considered overall poor quality
 COMMENT possible reversed clone: similarity on wrong strand
 Seq primer: -40RP from Gibco
 High quality sequence stop: 1.
 Location/Qualifiers
 1..214

FEATURES
 source

/organism="Meloidogyne incognita"
 /db_xref="taxon:6306"
 /clone_lib="Bird-Rao Meloidogyne incognita J2"
 /dev_stage="enriched for 2nd stage juveniles"
 /lab_host="X10RL"
 /note="Vector: ZAP express - pBCKW (Stratagene); Site_1: EcoRI; Site_2: XhoI; Oligo (dT) primed library. cDNA was constructed and cloned unidirectionally into the vector within the 5' EcoRI and 3' XhoI sites. This library was constructed by Dr. Uma Rao and Dr. David Bird at North Carolina State University."

BASE COUNT 59 a 52 c 19 g 84 t
 ORIGIN

Query Match 60.0%; Score 15; DB 121; Length 214;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 caagtcacagaatta 22
|||||
Db 184 CAACTCACAGAAATTA 170

RESULT 28
BE763893 233 bp mRNA EST 19-SEP-2000
LOCUS RC4-NT0054-120600-015-g05 NT0054 Homo sapiens cDNA, mRNA sequence.
DEFINITION BE763893
ACCESSION BE763893
VERSION BE763893.1 GI:10193817
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 233)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,M. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
COMMENT Contact: Simpson A.J.J.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=612-RC4-NT0054-120
600-015-g05&t3=2000-06-12&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 233.
Location/Qualifiers
1. 233
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="NT0054"
/dev_stage="Adult"
/note="Organ: nervous_tumor; Vector: puc18; Site_1: Sma1;
Site_2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 88 a 27 c 52 g 66 t
ORIGIN

Query Match 60.0%; Score 15; DB 139; Length 233;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 caagtcacagaatta 22
|||||
Db 94 CAACTCACAGAAATTA 108

RESULT 29
BE763892

LOCUS BE763892 253 bp mRNA EST 19-SEP-2000
DEFINITION RC4-NT0054-120600-015-g03 NT0054 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE763892
VERSION BE763892.1 GI:10193816
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 253)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,M. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
COMMENT Contact: Simpson A.J.J.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=612-RC4-NT0054-120
600-015-g03&t3=2000-06-12&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 12
High quality sequence stop: 253.
Location/Qualifiers
1. 253
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="NT0054"
/dev_stage="Adult"
/note="Organ: nervous_tumor; Vector: puc18; Site_1: Sma1;
Site_2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 95 a 31 c 56 g 71 t
ORIGIN

Query Match 60.0%; Score 15; DB 139; Length 253;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 caagtcacagaatta 22
|||||
Db 114 CAACTCACAGAAATTA 128

RESULT 30
AA317110 256 bp mRNA EST 19-APR-1997
LOCUS AA317110
DEFINITION EST19001 Lung Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION AA317110
VERSION AA317110.1 GI:1969448
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 256)

AUTHORS
 Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bull
 C.J., Lee,N.H., Kirkness,E.F., Weinstein,K.G., Gocayne,J.D., White
 O., Sutton,G., Blake,J.A., Brandon,R.C., Man-ai,C., Clayton,R.A.,
 Clive,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
 L.M., Fitzhugh,W.M., Fitchman,J.L., Geoghagen,N.S., Glodde,A.,
 Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
 Moreno-Palanges,R.F., McDonald,L.A., Nguyen,D.T., Pellegrino,S.M.,
 Phillips,C.A., Ryder,S.E., Scott,J.L., Saudak,D.M., Shirley,R.,
 Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
 Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
 Dimke,D., Feng,D.-F., Ferris,A., Fischer,C., Hastings,G.A., He,M.W.,
 Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
 Kunsch,C., Hungjun,J., Li,H., Meisner,P.S., Olsen,H., Raymond,L.,
 Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Pannon
 M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
 Venter,J.C.

TITLE
 Initial assessment of human gene diversity and expression patterns
 based upon 83 million nucleotides of cDNA sequence

JOURNAL
 Nature 377 (6547 Suppl), 3-174 (1995)

COMMENT
 Other_ESTS: TNC99418
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please check the TIGR Human Gene
 Index (<http://www.tigr.org/tldb/hgi/hgi.html>)
 Seq primer: M13 Reverse

FEATURES
 Location/Qualifiers
 1..256
 /organism="Homo sapiens"
 /db_xref="ATCC (lnhost):117571"
 /db_xref="taxon:9606"
 /clone_lib="lung"
 /dev_stage="adult"
 /note="Organ: Lung; Vector: pBluescript SK-; Site_1: EcoRI
 ; Site_2: XhoI"

BASE COUNT
 73 a 57 c 48 g 75 t 3 others

ORIGIN

Query Match 60.0%; Score 15; DB 5; Length 256;
 Best Local Similarity 100.0%; Pred. No. 98;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 caagtcacagaatta 22
 |||
 Db 189 CAAGTCACAGATT 203

RESULT 31
 245844 261 bp mRNA EST 14-NOV-1994
 LOCUS HSC2VE041 normalized infant brain cDNA Homo sapiens cDNA clone
 DEFINITION c-zve04, mRNA sequence.
 ACCESSION 245844
 VERSION 245844.1 GI:575078
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 261)
 Aufray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes
 M.D., Duprat,S., Hougaite,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
 Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
 Sebastiant-Kabakchis,C. and Tessier,A.
IMAGE: molecular integration of the analysis of the human genome

JOURNAL
 C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)

COMMENT
 and its expression
 Contact: Genethon
 Genethon Centre de recherche sur le Genome Humain
 1, rue de l'Internationale, BP60 91002 Evry Cedex, FRANCE
 Tel: 33169472800
 Fax: 33160778698
 Email: genexpress@genethon.fr
 Single read
 Genexpress_Library_id: C; Genexpress_sequence_id: y1c-zve04
 Seq primer: (-21)M13 universal

FEATURES
 Location/Qualifiers
 1..261
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="normalized infant brain cDNA"
 /sex="Female"
 /tissue_type="total brain"
 /dev_stage="3 months old"
 /note="Organ: brain; Vector: lambda BA; Site_1: HindIII;
 Site_2: NotI, sex=Female; dev_stage=3 months old;
 Isolate=muscular atrophy patient; tissue_type=total brain
 ; total mRNA was oligo-(dT) primed and directionally
 cloned 5' -> 3' into the HindIII -> NotI sites of the
 lambda BA vector. Clone library from B.Souares, Psychiatry
 Dept. Columbia University, USA. Normalization_method:
 Bento Soares, P.N.A.S. in press"

BASE COUNT
 77 a 56 c 31 g 97 t

ORIGIN

Query Match 60.0%; Score 15; DB 191; Length 261;
 Best Local Similarity 100.0%; Pred. No. 98;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 caagtcacagaatta 22
 |||
 Db 138 CAAGTCACAGATT 124

RESULT 32
 BF745516 263 bp mRNA EST 10-JAN-2001
 LOCUS CM2-BT0826-021100-469-h05 BT0826 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF745516
 VERSION BF745516.1 GI:12072192
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 263)
 Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Brites,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
**Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags**
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

BASE COUNT 68 a 75 c 67 g 72 t
 ORIGIN

Query Match 60.0%; Score 15; DB 169; Length 282;
 Best Local Similarity 100.0%; Pred. No. 99;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 attgtcgaatcac 15
 |||||||
 Db 51 ATTGTCCAGATCAC 37

RESULT 35
 AA344205 286 bp mRNA EST 21-APR-1997
 LOCUS EST50086 Gall bladder 1 Homo sapiens cDNA 5' end, mRNA sequence.
 DEFINITION
 ACCESSION AA344205
 VERSION AA344205.1 GI:1996443
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 286)
 Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult
 C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White
 O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Val, C., Clayton, R.A.,
 Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald
 L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghegan, N.S., Glodok, A.,
 Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,
 Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,
 Moreno-Palancas, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,
 Phillips, C.A., Ryder, S.E., Scott, J.L., Saudak, D.M., Shirley, R.,
 Small, R.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,
 Bedarick, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
 Dimke, D., Feng, D.-F., Ferlie, A., Fischer, C., Hastings, G.A., He, W.W.,
 Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,
 Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H., Raymond, L.,
 Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon
 M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and
 Venter, J.C.

TITLE Initial assessment of human gene diversity and expression patterns
 based upon 83 million nucleotides of cDNA sequence
 JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
 MEDLINE 96026280
 COMMENT Other ESTs: THG99418
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org

For clone availability, additional sequence and expression
 information related to this EST, please check the TIGR Human Gene
 Index (<http://www.tigr.org/tdb/hg1/hg1.html>)
 Seq primer: M13 Reverse.

FEATURES
 source Location/Qualifiers

1..286
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):145738"
 /db_xref="taxon:9606"
 /clone_lib="Gall bladder I"
 /sex="female"
 /dev_stage="adult, 25 yrs"
 /note="Organ: gall bladder; Vector: pBluescript SK-";
 Site_1: EcoRI; Site_2: XhoI
 BASE COUNT 83 a 62 c 48 g 91 t 2 others
 ORIGIN

Query Match 60.0%; Score 15; DB 6; Length 286;
 Best Local Similarity 100.0%; Pred. No. 99;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 caagtcaagaatta 22
 |||||||
 Db 197 CAAGTCACAGATTA 211

RESULT 36
 B53227 292 bp DNA GSS 20-JUN-1998
 LOCUS CIT-HSP-2009H2.TR CIT-HSP Homo sapiens genomic clone 2009H2, DNA
 DEFINITION
 ACCESSION B53227
 VERSION B53227.1 GI:2607561
 KEYWORDS
 SOURCE GSS.
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 292)
 Adams, M.D., Rounsley, S.D., Field, C.E., Bass, S., Linher, K., Golden
 K., Berry, K., Rounsley, S.D., Suh, E., Wible, C., Shizuya, H., Simon, M.
 and Venter, J.C.

TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
 Building
 JOURNAL Unpublished (1997)
 COMMENT Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC
 end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
 Seq primer: M13 Reverse
 Class: BAC ends.

FEATURES
 source Location/Qualifiers

1..292
 /organism="Homo sapiens"
 /db_xref="GDB:7040812"
 /db_xref="taxon:9606"
 /clone="2009H2"
 /clone_lib="CIT-HSP"
 /sex="Male"
 /cell_type="Sperm"
 /note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
 HindIII"
 BASE COUNT 58 a 64 c 90 g 80 t
 ORIGIN

Query Match 60.0%; Score 15; DB 256; Length 292;
 Best Local Similarity 100.0%; Pred. No. 99;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 caagtcaagaatta 22
 |||||||
 Db 248 CAAGTCACAGATTA 234

RESULT 37
 BF743951 317 bp mRNA EST 10-JAN-2001
 LOCUS BF743951
 DEFINITION CM2-BT0826-181000-428-c02 BT0826 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF743951
 VERSION BF743951.1 GI:12070627
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 317)

DIAS NETO, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL MEDLINE 20202663

TITLE Initial Assessment of Human Gene Diversity and Expression Patterns Based Upon 83 Million Basepairs of cDNA Sequence

COMMENT Nature 377, 3-174 (1995)

Other ESTs: THC10632

Contact: Venter, J.C.

The Institute for Genomic Research

932 Clopper Rd, Gaithersburg, MD 20878

Tel: 3018699056

Fax: 3018699423

Email: tdbinfo@tdb.tigr.org

For clone availability, additional sequence and expression information related to this EST, please contact the TIGR Database (tdbinfo@tdb.tigr.org)

Seq primer: M13 Reverse.

FEATURES

source Location/Qualifiers

1..317

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="Human Brain"

/note="Organ: brain"

BASE COUNT 89 a 75 c 47 g 115 t 1 others

ORIGIN

Query Match 60.0%; Score 15; DB 188; Length 327;

Best Local Similarity 100.0%; Pred. No. 99;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 caagtcacagaatta 22

DB 295 CAACTCAGCAATTA 281

LOCUS BE743964 335 bp mRNA EST 10-JAN-2001

DEFINITION CM2-BT0826-181000-428-H04 BT0826 Homo sapiens cDNA, mRNA sequence.

ACCESSION BF743964

VERSION BF743964.1 GI:12070640

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 335)

DIAS NETO, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL MEDLINE 20202663

TITLE Initial Assessment of Human Gene Diversity and Expression Patterns Based Upon 83 Million Basepairs of cDNA Sequence

COMMENT Nature 377, 3-174 (1995)

Other ESTs: THC10632

Contact: Venter, J.C.

The Institute for Genomic Research

932 Clopper Rd, Gaithersburg, MD 20878

Tel: 3018699056

Fax: 3018699423

Email: tdbinfo@tdb.tigr.org

For clone availability, additional sequence and expression information related to this EST, please contact the TIGR Database (tdbinfo@tdb.tigr.org)

Seq primer: M13 Reverse.

FEATURES

source Location/Qualifiers

1..327

/organism="Homo sapiens"

/db_xref="ATCC (inhost):10082"

/db_xref="taxon:9606"

/clone_lib="Human Brain"

/note="Organ: brain"

BASE COUNT 89 a 75 c 47 g 115 t 1 others

ORIGIN

Query Match 60.0%; Score 15; DB 169; Length 317;

Best Local Similarity 100.0%; Pred. No. 99;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gcaagtcacagaatt 21

DB 45 GCAAGTCACAGATT 31

LOCUS T30901 327 bp mRNA EST 06-SEP-1995

DEFINITION EST24536 Human Brain Homo sapiens cDNA 5' end similar to None, mRNA sequence.

ACCESSION T30901

VERSION T30901.1 GI:612999

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 327)

ADAMS, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Chiu, M.W., Clayton, R.A., Cline, R.T., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghegan, N.S.M., Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, J.P., Kelley, J.M.,

Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/ICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?Pl1-CM2&t2-CM2-BT0826-181000-428-h04&t3=2000-10-18&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 40
 High quality sequence stop: 335.
 Location/Qualifiers

FEATURES
 source
 1..335
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="BT0826"
 /dev_stage="Adult"
 /note="Organ: breast; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 /716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT
 101 a 76 c 55 g 103 t

Query Match
 Best Local Similarity 60.0%; Score 15; DB 169; Length 335;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 7 gcaagtcacagaatt 21
 |||||||||||||
 Db 61 GCAAGTCACAGAATT 47

RESULT 40
 BF743954/C 346 bp mRNA EST 10-JAN-2001
 LOCUS CM2-BT0826-181000-428-d07 BT0826 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BF743954
 ACCESSION BF743954.1 GI:12070630
 VERSION
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 (bases 1 to 346)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.G.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/ICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?Pl1-CM2&t2-CM2-BT0826-181000-428-d07&t3=2000-10-18&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 43
 High quality sequence stop: 346.

FEATURES
 source
 Location/Qualifiers

1..346
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="BT0826"
 /dev_stage="Adult"
 /note="Organ: breast; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 /716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT
 103 a 78 c 54 g 111 t

Query Match
 Best Local Similarity 60.0%; Score 15; DB 169; Length 346;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 7 gcaagtcacagaatt 21
 |||||||||||||
 Db 73 GCAAGTCACAGAATT 59

RESULT 41
 T07051 370 bp mRNA EST 30-JUN-1993
 LOCUS EST04940 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA
 DEFINITION clone HFBEC79, mRNA sequence.
 ACCESSION T07051
 VERSION T07051.1 GI:318200
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 (bases 1 to 370)
 Adams, M.D., Kerlavage, A.R., Fields, C. and Venter, J.C.
 3,400 expressed sequence tags identify diversity of transcripts from human brain
 Nature Genet. 4, 256-267 (1993)
 9364420
 JOURNAL MEDLINE
 COMMENT Contact: Adams, MD
 The Institute for Genomic Research
 932 Clopper Road, Gaithersburg, MD 20878
 Tel: 3018699056
 Fax: 3018699423
 Email: mdadams@ligr.org
 Seq primer: M13-21.

FEATURES
 source
 Location/Qualifiers

1..370
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):83698"
 /db_xref="taxon:9606"
 /clone_lib="HFBEC79"
 /note="Organ: Fetal brain, Stratagene (cat#936206)"
 /note="Vector: LambdaZAP-II; 17-18 wk gestation, female; oligo-dT + random primed cDNA synthesis; lambdaZAP-II vector, 1.0kb average insert size."
 BASE COUNT
 137 a 58 c 68 g 106 t 1 others

Query Match
 Best Local Similarity 60.0%; Score 15; DB 188; Length 370;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 8 caagtcacagaatta 22
 |||||||||||||
 Db 113 CAAGTCACAGAATTA 127

```

RESULT 42
H53286/c 381 bp mRNA EST 20-SEP-1995
LOCUS Y983109.r1 Soares fetal liver spleen 1NFUS Homo sapiens cDNA clone
DEFINITION IMAGE:202601.5', mRNA sequence.
ACCESSION H53286
VERSION H53286.1 GI:993433
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 381)
AUTHORS Hillier,L., Clark,N., Dnubque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Maita,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevisakis,E., Waterston
,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1459
High quality sequence stops: 288
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1459 Std Error: 0.00
Seq primer: M13Rpi
High quality sequence stop: 288.

FEATURES
source
1. .381
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:371732"
/db_xref="taxon:9606"
/clone="IMAGE:202601"
/clone_1lb="Soares fetal liver spleen 1NFUS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: p773D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACGTGAGAGATTTAATAAGATCTTTTCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 105 a 99 c 79 g 88 t 10 others
ORIGIN
Query Match 60.0%; Score 15; DB 158; Length 381;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 gtacacgaattatt 25
|||||
Db 281 gtccacgaattatt 267

```

```

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 395)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui, M.D.
, Michael Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/bdhp/image/image.html
Insert Length: 622 Std Error: 0.00
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 370.

FEATURES
source
1. .395
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1011676"
/clone_1lb="NCI-CGAP_P13"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from 5,000-10
,000 microdissected cells histologically-determined to be
fully malignant prostate cancer cells. Double-stranded
cDNA was ligated to EcoRI adaptors, 5 cycles of PCR
applied to the cDNA with an adaptor-specific primer, and
the resulting PCR product subcloned into pAMP10 by the
UDG-cloning method (Life Technologies). Average insert
size is 600 bp. NOTE: Not directionally cloned. This
library was constructed by David Krizman."

BASE COUNT 110 a 62 c 65 g 158 t
ORIGIN
Query Match 60.0%; Score 15; DB 4; Length 395;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 aagtcacgaattat 23
|||||
Db 83 AAGTCACGAAATTAT 69

RESULT 44
A1694429 402 bp mRNA EST 02-JUN-1999
LOCUS wd83f06.x1 NCI-CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2338211 3',
DEFINITION mRNA sequence.
ACCESSION A1694429
VERSION A1694429.1 GI:4971769
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 402)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)

```

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bdrip/image/image.html
 Seq primer: -40UP from Gibco.
 Location/Qualifiers

FEATURES

source

1..402
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2338211"
 /clone_lib="NCI-CGAP_Lu24"
 /tissue_type="carcinoid"
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI-CGAP_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

160 a 60 c 69 g 113 t

ORIGIN

Query Match 60.0%; Score 15; DB 23; Length 402;
 Best Local Similarity 100.0%; Pred. No. 99;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 caagtcacagaattta 22
 |||

Db 342 CAAGTCACAGAATTa 356

RESULT 45

AI203847

LOCUS

qf75h03.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1755893

DEFINITION

3', mRNA sequence.

ACCESSION

AI203847
 AI203847.1 GI:3756453

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 404)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgabs-r@mail.nih.gov

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo

, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bdrip/image/image.html

Insert Length: 563 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 394.

Location/Qualifiers

1..404

/organism="Homo sapiens"

/db_xref="taxon:9606"
 /clone="IMAGE:1755893"
 /clone_lib="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dt) primer [5' TGTTACCAATCTGTAAGGCGAGCGCCGCCAATTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

161 a 61 c 69 g 113 t

ORIGIN

Query Match 60.0%; Score 15; DB 17; Length 404;
 Best Local Similarity 100.0%; Pred. No. 99;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

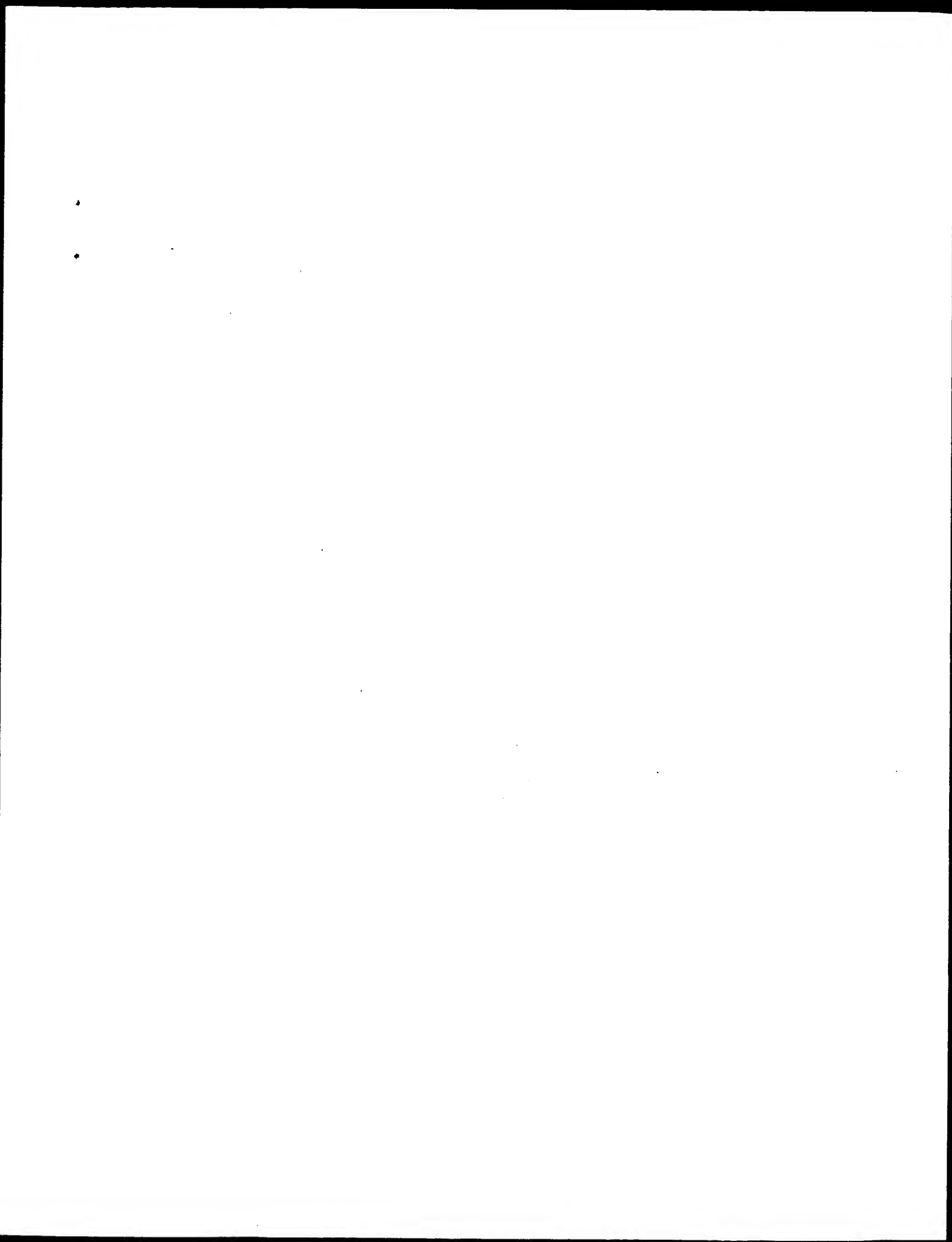
QY 8 caagtcacagaattta 22
 |||

Db 342 CAAGTCACAGAATTa 356

Search completed: October 9, 2001, 15:15:23
 Job time: 13645 sec

Wed Oct 10 07:45:42 2001

us-09-396-196f-2.01i.rst



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:44:37 ; Search time 1670.83 Seconds

(without alignments)
231.438 Million cell updates/sec

Title: US-09-396-196f-3

Sequence: 1 gtgtctgacatctgtgttgaagcg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 segs, 773874588 residues

Word size : 9

Total number of hits satisfying chosen parameters: 254803

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:

1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vil:*
59: gb_viz:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_pr10:*
95: gb_pr11:*
96: gb_pr12:*
97: gb_pr13:*
98: em_ba3:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	25	100.0	128	2	ECOBIOB	M27731 Escherichia
2	25	100.0	1041	9	AR029499	AR029499 Sequence
3	25	100.0	1041	9	AR034916	AR034916 Sequence
4	25	100.0	1084	9	A11530	A11530 B10b gene o
5	25	100.0	5526	2	AF250776	AF250776 Uncultured
6	25	100.0	5793	2	ECOBIOB	J04423 E.coli 7,8-
7	25	100.0	5872	9	A38246	A38246 Sequence 1
8	25	100.0	5872	9	A38251	A38251 Sequence 6

```

9      25      100.0      5872      9      A93674      A93674 Sequence 1
10     25      100.0      5872      9      A93679      A93679 Sequence 6
11     25      100.0      5872      9      AR101809      AR101809 Sequence
12     25      100.0      5872      9      AR101810      AR101810 Sequence
13     25      100.0      11022      1      AE000180      AE000180 Escherich
14     25      100.0      13501      1      AE005258      AE005258 Escherich
15     25      100.0      297816      2      AP002553      AP002553 Escherich
16     24      96.0      1121      10      E00893      E00893 Genomic DNA
17     20      80.0      12891      1      AE004192      AE004192 Vibrio ch
18     17      68.0      383      96      S64534      S64534 Drosophila
19     17      68.0      461      96      S64532      S64532 Drosophila
20     17      68.0      516      6      DMDECG1B      X76231 D.melanog
21     17      68.0      539      96      S64530      S64530 Drosophila
22     17      68.0      539      96      S64531      S64531 Drosophila
23     17      68.0      617      96      S64529      S64529 Drosophila
24     17      68.0      774      6      DSDECTTR6      X96931 D.secellia
25     17      68.0      852      6      DSDECTTR2      X96927 D.simulans
26     17      68.0      852      6      DSDECTTR3      X96928 D.simulans
27     17      68.0      930      6      DMDECTTR1      X96926 D.melanog
28     17      68.0      3741      6      DMDECTTR1      M35887 Drosophila
29     17      68.0      3755      6      DMDECTTR1      M35888 Drosophila
30     17      68.0      5697      6      DMDECTTR1      M35889 Drosophila
31     17      68.0      33204      65      AC017995      AC017995 Drosophila
32     17      68.0      161540      64      AC016952      AC016952 Homo sapi
33     17      68.0      306267      4      AE003442      AE003442 Homo sapi
34     16      64.0      774      6      DMDECTTR5      X96930 D.mauritian
35     16      64.0      1614      8      AF281080      AF281080 Xenopus l
36     16      64.0      5000      13      AF143940      AF143940 Arabidops
37     16      64.0      6887      15      Y1012084      Y1012084 Arabidops
38     16      64.0      11328      1      AE004322      AE004322 Vibrio ch
39     16      64.0      13856      1      AE004026      AE004026 Xylella f
40     16      64.0      15247      1      AE004580      AE004580 Pseudomon
41     16      64.0      31557      92      HS314G4      Z69667 Human DNA s
42     16      64.0      41200      14      SC8520X      Z49705 S.cerevisia
43     16      64.0      46340      9      AX008043      AX008043 Sequence
44     16      64.0      62927      12      AB018119      AB018119 Arabidops
45     16      64.0      63679      63      AC013956      AC013956 Drosophila

```

ALIGNMENTS

```

RESULT 1
ECOBIOB      128 bp      DNA      BCT      03-JAN-1995
LOCUS      Escherichia coli biotin (biob) gene, early terminator region.
DEFINITION      M27731
VERSION      M27731.1 GI:341755
KEYWORDS      biob gene; biotin.
SOURCE      Escherichia coli (strain K-12) DNA.
ORGANISM      Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE      1 (bases 1 to 128)
AUTHORS      Nath,S.K.
TITLE      Attenuation of transcription of biotin genes in Escherichia coli
JOURNAL      Can. J. Microbiol. 34 (12), 1288-1296 (1988)
MEDLINE      89167942
FEATURES
    source          location/Qualifiers
        gene         1..128
                    /organism="Escherichia coli"
                    /strain="K-12"
                    /db_xref="taxon:562"
        -10_signal   4..128
                    /gene="biob"
        gene         4..9
                    /gene="biob"
        mRNA         20..>128
                    /gene="biob"
        terminator   103..121
                    /gene="biob"
                    /note="early terminator"
BASE COUNT      31 a      28 c      34 g      35 t

```

ORIGIN

```

Query Match      100.0%; Score 25; DB 2; Length 128;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1      gtgtcgatcgtgttgaagc 25
Db      18      GTTCTGATCTGCTGTTGAAGCG 42

```

RESULT 2

```

LOCUS      AR029499      1041 bp      DNA      PAT      29-SEP-1999
DEFINITION      Sequence 7 from patent US 5859335.
ACCESSION      AR029499
VERSION      AR029499.1 GI:5941472
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unknown.
REFERENCE      1 (bases 1 to 1041)
AUTHORS      Patton,D.Andrew.
TITLE      Enhanced biotin biosynthesis in plant tissue
JOURNAL      Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
    source          location/Qualifiers
        gene         1..1041
                    /organism="unknown"
BASE COUNT      262 a      273 c      305 g      201 t

```

```

Query Match      100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1      gtgtcgatcgtgttgaagc 25
Db      57      GTTCTGATCTGCTGTTGAAGCG 81

```

RESULT 3

```

LOCUS      AR034916      1041 bp      DNA      PAT      29-SEP-1999
DEFINITION      Sequence 7 from patent US 5869719.
ACCESSION      AR034916
VERSION      AR034916.1 GI:5950521
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unknown.
REFERENCE      1 (bases 1 to 1041)
AUTHORS      Patton,D.A.
TITLE      Transgenic plants having increased biotin content
JOURNAL      Patent: US 5869719-A 7 09-FEB-1999;
MEDLINE
FEATURES
    source          location/Qualifiers
        gene         1..1041
                    /organism="unknown"
BASE COUNT      262 a      273 c      305 g      201 t

```

```

Query Match      100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1      gtgtcgatcgtgttgaagc 25
Db      57      GTTCTGATCTGCTGTTGAAGCG 81

```

RESULT 4

A11530
 LOCUS A11530 1084 bp DNA PAT 09-FEB-1994
 DEFINITION BioB gene of E.coli with primers.
 ACCESSION A11530
 VERSION A11530.1 GI:490218
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
 REFERENCE 1 (bases 1 to 1084)
 AUTHORS
 JOURNAL Patent: GB 2216530-A 16-11-OCT-1989;
 FEATURES
 source
 Location/Qualifiers
 1..1084
 /organism="Escherichia coli"
 /db_xref="taxon:562"
 24..1064
 /gene="BIOB"
 24..1064
 /gene="BIOB"
 /codon_start=1
 /transl_table=11
 /protein_id="CAA00965.1"
 /db_xref="GI:490219"
 /db_xref="SWISS-PROT:P12996"
 /translation="MAHRPMTLSOVTELEFEKPLDLLLFEAOQVHRQHPDPROVYST
 LLSIKTGACPEDECKYCPOSSRYKTGLEARLMEVOYLESARAKAAGSTRPCGAAM
 KNPNERDMPYLEOMVOGVKAMGLEACMTIGTISEQAORLANAGLDYYVNHNLDTSPER
 YGNLITTTTORYERLDITLEKVRDAGIKVCSGGIVGSEYKDRAGLLLOANLPTPES
 VPIMLVVKVGTPLADNDVDAPDFIRITAVARIMPTSYVLSAGREOMNORQAMC
 PMAGANSIFYGCKLITTTNPREDKDLOLFRKGLINPQOTAVLAGDNEQORLEQALMT
 PTDTERYNAAL"
 BASE COUNT 271 a 286 c 318 g 209 t
 ORIGIN
 Query Match 100.0%; Score 25; DB 9; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 0.0016;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 gtcgtgacatcgtgttgaagcg 25
 |||||||||||||||||||||
 Db 80 GTTCTGATCTGCTGTTTCAACG 104
 RESULT 5
 AF250776 BCT 31-JAN-2001
 LOCUS
 DEFINITION uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in
 modC-bioA intergenic region, DAPA-aminotransferase BioB (bioB),
 biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and
 biotin biosynthesis protein BioC (bioC) genes, complete cds; and
 dehydrobiotin synthetase BioD (bioD) gene, partial cds.
 ACCESSION AF250776
 VERSION AF250776.1 GI:12620124
 KEYWORDS
 SOURCE uncultured bacterium pCosHE2.
 ORGANISM Bacteria; environmental samples.
 REFERENCE 1 (bases 1 to 5526)
 AUTHORS Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streitt,W.R.
 TITLE Direct cloning from enrichment cultures, a reliable strategy for
 isolation of complete operons and genes from microbial consortia
 Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
 JOURNAL
 MEDLINE 11133432
 PUBMED 20575196
 REFERENCE 2 (bases 1 to 5526)
 AUTHORS Entcheva,P., Liebl,W. and Streitt,W.R.
 TITLE Direct Submission
 JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
 Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
 FEATURES
 Location/Qualifiers

source
 1..5526
 /organism="uncultured bacterium pCosHE2"
 /db_xref="taxon:143797"
 /clone="pCosHE2"
 /note="unknown organism, cosmid clone derived from
 environmental consortium"
 complement(52..528)
 /note="ORF1"
 /codon_start=1
 /transl_table=11
 /product="hypothetical 17.1 kDa protein in modC-bioA
 intergenic region"
 /protein_id="AAG60577.1"
 /db_xref="GI:12620125"
 /translation="MKLISNDLRDQDKLPHRHVYNGGYGDNISPLHADDYPACTK
 SFVYTCYDPDAPGSGMMHWVNLPAIDTRYLPQGFSGGLVAMPDGVLTQRTDPRKAG
 YDGAAPPGETHRIFVYHALDIRIDVDEGASGAMGFVNHSLASASITAMFS"
 complement(587..1876)
 /gene="bioA"
 complement(587..1876)
 /note="7,8-diaminopelargonic acid
 synthetase-aminotransferase"
 /codon_start=1
 /transl_table=11
 /product="DAPA-aminotransferase BioA"
 /protein_id="AAG60578.1"
 /db_xref="GI:12620126"
 /translation="MTDDLLFDDRHIMPTSMTPPLPYVVSABGCELLISDGR
 LVDSGSMMAIHYVNPOLNAKMSQIDAMSHVFGIITHAPAEICRKLVMTPQP
 LKCYFLADSGVAEVAAMKALQWQAKGROEFLFRNGYHGTDTFGAMSDYDGS
 MSLMLKGYLPENLPAPOSMDGEMDGPVAPARLMAARHRIITAVIIEPYDGS
 GMRTHPEMLKRIKICDREGILILADIATNGFRTKTRACAEALTAIPOLICGRKL
 TGGWTTSATLTTRVAVETISNGAGCPMHPGPNPLACAAANSLADESDWQ
 OVADIEVOLREQLAPARDAEMVADVRLVGAIGVETHPVMAALQFVEQGVWIRP
 FGKLITLMPYIILPOOLRLTAVNAVODEPFCQ"
 1963..3003
 /gene="bioB"
 1963..3003
 /gene="bioB"
 /codon_start=1
 /transl_table=11
 /product="biotin synthase BioB"
 /protein_id="AAG60579.1"
 /db_xref="GI:12620127"
 /translation="MAHRPMTLSOVTELEFEKPLDLLLFEAOQVHRQHPDPROVYST
 LLSIKTGACPEDECKYCPOSSRYKTGLEARLMEVOYLESARAKAAGSTRPCGAAM
 KNPNERDMPYLEOMVOGVKAMGLEACMTIGTISEQAORLANAGLDYYVNHNLDTSPER
 YGNLITTTORYERLDITLEKVRDAGIKVCSGGIVGSEYKDRAGLLLOANLPTPES
 VPIMLVVKVGTPLADNDVDAPDFIRITAVARIMPTSYVLSAGREOMNORQAMC
 PMAGANSIFYGCKLITTTNPREDKDLOLFRKGLINPQOTAVLAGDNEQORLEQALMT
 PTDTERYNAAL"
 3000..4154
 /gene="bioF"
 3000..4154
 /gene="bioF"
 /note="8-amino-7-oxononanoate synthase"
 /codon_start=1
 /transl_table=11
 /product="KAPA synthetase BioF"
 /protein_id="AAG60580.1"
 /db_xref="GI:12620128"
 /translation="MSWQEKIINALDARRADALRRRPVPAQAGRNLVADDRYLNF
 SSNDYLGSHHPQILIRAOQGADEPGTSGSGSHVSGSVVHQALEELBLWSTSLRA
 LLFISGFANQAVIAAMAKEDRIADRLSHASLSLEASLSPSQLRFAHNDVHLAR
 LIASPCPOOLVTEGVFSMDGDSAPLAEIOGVQOHNGLMADVDAHGTGTGCGGSG
 SCMLQKVELLVVTEFGKFGVSGAAVCSSTVADYILQFARHLIYSTMPADAOAL
 RASLAVIRDEGDARREKLVSLIAHFRAGVODLPFTLADSCSAIOLPIYVGNSSALDL
 AEKLQGCWYTAIRPPTVPAGTARLRLITIAHMODIDLLEVLHNG"
 4141..4896
 /gene="bioC"
 4141..4896
 /gene="bioC"
 /gene="bioC"

/protein_id="AAA23518.1"
/db_xref="GI:145428"
/translation="MSKRYFVGTDTGKCTVASCALLAAKARAGYRTAGYKPVASGS
ETTPGELRNSDALAQORNSLDDATVNPYFAEPTSPHISAOGRPIESLVASG
LRLAEQADWVLEVGAGWETPLSDTEFFADWVQPOLPVLVGVKLCCLNHMLTA
QVIOHAGLTLTLAGWVANDVTPGKRHAEEVMTILRMI PRRCWERSPLGIKIQMRQES
T"

BASE COUNT 1363 a 1554 g 1631 q 1245 t
ORIGIN 4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match 100.0%; Score 25; DB 2; Length 5793;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgcggaactgcctgttgaaagc 25
Db 2068 GTTGCTGATCTCTGTTGAAGCG 2092
IIIIIIIIIIIIIIIIIIIIIIIIIIIIII

RESULT 7
LOCUS A38246 5872 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 1 from Patent WO9408023.
ACCESSION A38246
VERSION A38246.1 GI:2294844
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL Patent: WO 9408023-A 1 14-APR-1994;
LONZA AG (CH)
COMMENT other publication PL 308301 950724
other publication CA 2145400 940414
other publication AU 4820293 940426
other publication HU 71781 960228
other publication SK 42095 951108
other publication CZ 9500809 950913
other publication FI 951547 950331
other publication JP 85016947 960227.
FEATURES
source location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A-15/9"
1..96
/function="PROMOTER PTAC"
/evidence=experimental
23..28
45..49
/standard_name="PROMOTER PTAC"
/evidence=experimental
105..119
/standard_name="BIOB RBS NO.9"
/evidence=experimental
117..1157
/gene="BIOB"
117..1157
/gene="BIOB"
/codon_start=1
/transl_table=11
/number=1
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAA02325.1"
/db_xref="GI:2294845"
/translation="MAHRPRMTLSQVTELFERPLDLLFEAQOVHROHDFPROVQST
LTIKTGACPEDCKYCPQSSRYKTGLFAERLMEVYQVLESARKAKAAGSTRFCMAAM

KNPHRDPVTECKVQKAMGLEACMTIGTILSESORARLANGLDYNNHNDTSPF
YGNITTRTYOERLDTLEKVRDAGIKVCSGIVGLGTVYDRAGLLQLLANLETPPS
VPIMLKVKCTPLADNDVAFDFITFIARIMPTSVRSAGREQNEEDQAMC
FMAAGANSLFYGCCKLITTPNPEKDLQFLRLGIMPQATVLAGDNEQDQRLQALMT
PDTEDEYNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/number=3
/product="PROTEIN"
/protein_id="CAA02326.1"
/db_xref="GI:2294846"
/translation="MATVKNQALIAAFGRAAHYEHOADLOROSADALLAMLPOKRYT
HVLADGGPGMSRHRERBAQVYALDSPMLVQAKRQADAHYLAGDIESIPLATA
TFDLAMSLAYOKGNSLALRELRYVRRKGYAFTLYOGSLPELHQAMQVDERP
HANRFLPDELDGSLNGVHVOHITQPTLWDDALSMRSLKIGATHLEGRDRL
TRSQLORLQALMPQOGKRYPLTYHFLGVIARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:
8-AMINO-7-OXONONANOATE AMINOTRANSF. "
/ec_number="2.6.1.62"
/codon_start=1
/transl_table=11
/number=5
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAA02327.1"
/db_xref="GI:2294847"
/translation="MTTDDLAFFDGRHIMPTYSPLPYVSVASGECCELLISDGR
LVDMSSWMAAHNGVNHNPOLNAAMKSOIDAMSHVFGITHAPAIELGRIVMTDOP
LECVFLADSGSAVEAVAKMALQYQAKGARQFLTFRRNGYHGDTEFGASVCDPRS
MHSIMKGYLPENLFPAPQOSRMDGEMDERMVGARLMAAHRHEIAVITEPIVOGAG
GMRVYHPBWLKRIRKICDREGILLIDEIATNGRRGKELPACHEAETALPCTICATL
TGGTMTLSATITTRVAVETISNGRAGCFPHNGPTFMGRPLCAANASLALIEGDMQO
QVADIEVQLEQALPARDAEVADVRLAIGVETTHPVMAALOKFEVGGVWIRP
FGKLIYLPYIILPQOLORTAIVNRAVODETFFCO"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/number=6
/evidence=experimental
/product="PROTEIN"
/protein_id="CAA02328.1"
/db_xref="GI:2294848"
/translation="MKLISNDLRDGDLPKRRHVNNGNGYCDNISPHLAWDDVPAGIK
SFVYTCIDPDAPGSGMMHWVYVNLPAIDTVLPOGSGSLVAMPDGLVLRTPFGKIG
YDGAAPPKGETHRIYFTVHALDIERIDVDGASGAMGVNHFHSLASASITAMFS"
5583..5605
5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
1318 a 1552 c 1695 g 1307 t

BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;

Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgctgctgttgaagcg 25
|||||
Db 173 GTTGTGATCTGCTTTGAAGCG 197

RESULT 8
LOCUS A38251 5872 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 6 from Patent WO9408023.
ACCESSION A38251
VERSION A38251.1 GI:2294849
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL Patent: WO 9408023-A 6 14-APR-1994;
LONZA AG (CH)

COMMENT
Other publication PL 308301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42095 951108
Other publication CZ 9500809 950913
Other publication FI 951547 950331
Other publication JP 8501694T 960227.
Location/Qualifiers
1. .5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141. .1156
/standard_name="BIOF RBS"
1154. .2308
/gene="BIOF"
1154. .2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/number=2
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAA02329.1"
/db_xref="GI:2294850"
/translation="MSWQEKINAALDARRAADALRRYPVAQAGRWLVADROYLNF
SSNDYLGISHRPOITRAWQOGEQFSGSGSHVSGYSVHQALEELAEMLQYSRA
LLFISGPANQAVIAAMAKEDRIADRLSHASLLEASLSPSOLRRFAHNDVTHLAR
LLASPCPGQGVVTEGVFSGMDGASAPLAIEIOVQOHNGMILMVDAGHGTGVIGSGRG
SCWLOKVPDELLVTFPGKGVSGSAAYVCSSTVADYLLQPARHLIYSSMPAQOAL
RASLAVIRSDBDARREKIALILTRFAGVDLPPTLLADSCSAIOLPLIVGNSTRLOL
AEKLRQGCWVAIRPPTVPAGTARLRLTLAAHEMODIDRLLELVHNG"

RBS
gene
CDS

3043. .3753
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="DTB SYNTHASE"

/protein_id="CAA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYPVNCTDTEVEKGTVAACALLOAAKAGYRTAGYKPVASGS
EKTPPEGLRNSDIALORNSLSQDITVNPYFAETSPHITSADGPRTEISLWSAG
LRALQDQADVLEAGAGMFTPLSDFTFADWVTOBOLPVILVQKLECIINHAHLTA
QVIOHAGLTLAGVANDVTPPGKRHAHEVMTTLRLMPLDLIELPWLAEENPENAAATGK
YINLAPEVASTLTGFTSRL"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgctgctgttgaagcg 25
|||||
Db 173 GTTGTGATCTGCTTTGAAGCG 197

RESULT 9
LOCUS A93674 5872 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 1 from Patent EP0798384.
ACCESSION A93674
VERSION A93674.1 GI:6741862
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O. and Brass, J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 1 01-OCT-1997;
LONZA AG (CH)

FEATURES
source Location/Qualifiers
1. .5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A-15/9"
1. .96
/function="PROMOTER PTAC"
/evidence=experimental
23. .28
/standard_name="PROMOTER PTAC"
105. .119
/evidence=experimental
117. .1157
/gene="BIOB"
117. .1157
/gene="BIOB"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHRPMTLSQVTELEFEKRLDLPFEAOVHROHDFPROGVST
LISIKTGACPEDCKYCPQSSRYKGLAEERLMEVEQLSARAKAAGSTRRCMGAM
KNHERMPYLRQMVQYKAMGLEACMTGLTSESOAORLANAGLDYVHNHIDTSPF
YGNITTRITQERLDLLEKRDGKIRVCSGCIYGESEYKDRAGLLIQLANLPTPES
VPINMLVKYGTPLANDVDADFTRITAVARIMPTSTVRLSAGREOMNRQTOHMC
FMAGANSIFYGCKILTTPNPEEDKDQLRKRIGLINDQOTAVIAGDNEODQRLQALMT
POTDEYYNAAL"
2284. .2297
/standard_name="BIOC RBS"
2295. .3050
/gene="BIOC"

-35_signal
-10_signal

RBS
gene
CDS

3043. .3753
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="DTB SYNTHASE"

```
CDS
2295..3050
/feature="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVKKOALIAAFGAAHRYOHADLOROSADALLAMIPKRYT
HYDAGCGGMSRRHREHRAQVYALDLSPPHVARQKADADHLAGDIESLPLATA
TIDLANSLAVQWCGHSTALRELKRVKGVAVFTLVGSLPELHQAQWAVLPER
HANRFLPDEIEQSLNGVYOHIIQIPTLMFDALSAMRSLGIGATHLHGRDRPRL
TRSOLORLQALAMPQOGRYPTLTHFLGVIARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/feature="BIOA"
/feature="BIOA"
/feature="BIOA"
/feature="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence="experimental"
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLAFDQRIHWPYTSPLPYPVSAEGCELLSDGR
LVDGMSWMAAIIHGYNHQOLNAAMKSOIDAMSHVFGGITPAALIELCKLVAAMP
LDCVFLADSGVAEVAEMKALQYQAKGEARQRELTENGXHGDTFGAMSCDDPP
MSLWKGVLPEMLFAPAPQSRMDGEMDERDVGAFARMAAHHEHTAAVITPEYOGAG
GMRNYPEMLKRIKICDEGILLIADETATGFGTKLFACEHAEIADIIICLKAL
TGSTWLSATLTITREVAETTSNGAGCEFMHGTENGNDPLACAAVAALIESGWQO
QVADLEVLQREQLAPARDAENVADVRLGALISVETTHPVNMAALOKFEVEGVWIRP
FGKLIYLPYIILPQOLQRLTAANRAVODETFEQ"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/feature="ORFI"
/feature="ORFI"
/feature="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence="experimental"
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDGDKLPHRHVNGMGYDGDNTSPHLAMDVPACTK
SFVVCYDDAPATGSGMWVYVNNLPADTRVLPGFGSLVAMPDGVLOTRDFGKTG
YDGAAPKKEITHRIYLVHADIERIDVDEGASGAMVGHNHFHSLASITAMFS"
5583..5605
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
1318 a 1552 c 1695 g 1307 t
BASE COUNT
ORIGIN
```

```
Query Match
Best Local Similarity 100.0%; Score 25; DB 9; Length 5872;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gttgctgactgctgcttgaagcg 25
|||||
Db 173 GTTGCTGATCTGCTTTGAACGC 197
```

```
RESULT 10
A93679 5872 bp DNA PAT 22-JAN-2000
LOCUS A93679
DEFINITION Sequence 6 from Patent EP0798384.
```

```
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE
ORGANISM Escherichia coli.
Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O. and Brass, J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 6 01-Oct-1997;
LONZA AG (CH)
FEATURES
source location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PRO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/feature="BIOF"
1154..2308
/feature="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence="experimental"
/product="DAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSWQEKINAALDARRAADALRRRYPVAGAGRWLVADROYLNF
SSNYLQIGSHHPQILRAMQCAEQFGISGSGSVQSVYHQALEBELAEWLYSRA
LIFTSGFANGAVIAAMAKEDRIADRLSHASILEASISPSQIRFAHNDVTHLAR
LLASPCPOQWVVEGVESMDGSAPIAEIOVQOHNWMLMVDADGTGCVIGQGRG
SCWLOKXPELLVYTPFGKFGVSGAAVLCSTVADYLLQFARHLIYSTMPAOAOL
RASLAVRSDGDAAREKLALITRFRAGVODDEPTLADSCSAIPIIYVNDNSALDL
AEKLRQGCWVTAIRPPTVPAIGARLRLTLTAHEMODIDRLLEVLHGNG"
```

```
RBS
3030..3045
/standard_name="BIOD RBS"
3043..3753
/feature="BIOD"
3043..3753
/feature="BIOD"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence="experimental"
/product="DTP SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYEYTGDFEYGVKTVASCAILLOAKAAGYRTAGKPYVSS
EKTEPGEKNSDALALQNSSLQDLYATVNYTPRAPERSPHISQGBRPLESYMSAG
LRALIEQADWYLVGAGGWFPLSDITTFADWYQEDLPVLYGVKLVGINHMLTA
QVIGHAGITLAGWVANDVTPPGKRHAEMTTLIRMLIPAPLIGELPWLAEPMENAATGK
YINLAFYDASTLIGTFSRL"
```

```
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN
```

```
Query Match
Best Local Similarity 100.0%; Score 25; DB 9; Length 5872;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 gttgctgactgctgcttgaagcg 25
|||||
Db 173 GTTGCTGATCTGCTTTGAACGC 197
```

```
RESULT 11
ARI01809
```

```

LOCUS       AR101809      5872 bp      DNA
DEFINITION   Sequence 1 from patent US 6083712.
ACCESSION    AR101809
VERSION      AR101809.1  GI:12812607
KEYWORDS     .
SOURCE       Unknown.
ORGANISM     Unknown.
REFERENCE    1 (bases 1 to 5872)
AUTHORS      Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE        Biotechnological method of producing biotin
JOURNAL      Patent: US 6083712-A 1 04-JUL-2000;
FEATURES     Location/Qualifiers
             source          1..5872
BASE COUNT   1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match          100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gttcgtgactcgtcgttgaagcg 25
    |||||||
Db 173 GTTCTGATCTGCTGTTGAAGCG 197

RESULT 12
LOCUS       AR101810      5872 bp      DNA
DEFINITION   Sequence 6 from patent US 6083712.
ACCESSION    AR101810
VERSION      AR101810.1  GI:12812608
KEYWORDS     .
SOURCE       Unknown.
ORGANISM     Unknown.
REFERENCE    1 (bases 1 to 5872)
AUTHORS      Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE        Biotechnological method of producing biotin
JOURNAL      Patent: US 6083712-A 6 04-JUL-2000;
FEATURES     Location/Qualifiers
             source          1..5872
BASE COUNT   1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match          100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gttcgtgactcgtcgttgaagcg 25
    |||||||
Db 173 GTTCTGATCTGCTGTTGAAGCG 197

RESULT 13
LOCUS       AE000180      11022 bp      DNA
DEFINITION   Escherichia coli K12 MG1655 section 70 of 400 of the complete genome.
ACCESSION    AE000180  U000096
VERSION      AE000180.1  GI:1786988
KEYWORDS     .
SOURCE       Escherichia coli K12.
ORGANISM     Escherichia coli K12.
REFERENCE    1 (bases 1 to 11022)
AUTHORS      Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Koe, C.K., Mayhew, G.F.,

```

```

Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.
The complete genome sequence of Escherichia coli K-12
Science 277 (5331), 1453-1474 (1997)
JOURNAL
MEDLINE
PUBMED
97426617
9278503
REFERENCE    2 (bases 1 to 11022)
AUTHORS      Blattner, F.R.
TITLE        Direct Submission
JOURNAL      Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
3 (bases 1 to 11022)
REFERENCE    3
AUTHORS      Blattner, F.R.
TITLE        Direct Submission
JOURNAL      Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
4 (bases 1 to 11022)
REFERENCE    4
AUTHORS      Plunkett, G. III.
TITLE        Direct Submission
JOURNAL      Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG site nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). ***The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES     Location/Qualifiers
             source          1..11022
             /organism="Escherichia coli K12"
             /strain="K12"
             /sub_strain="MG1655"
             /db_xref="taxon:83333"
             <1..130
             /note="REP (repetitive extragenic palindromic) element; contains 4 REP sequences"
             complement(147..1430)
             /gene="ybhC"
             /note="b0772"
             complement(147..1430)
             /gene="ybhC"
             /function="orf; NOT classified"
             /note="1427; 98 pct identical to fragment YBHC_ECOLI SW:P46130 (300 aa) but contains 127 additional C-terminal residues"
             /codon_start=1
             /transl_table=11
             /product="putative pectinesterase"

```

/protein_id="AAC73859.1"
 /db_xref="GI:1786989"
 /translation="MNTFSVRLALALAGVTLTAACSSPPDRPDRDOTAPCTSRPI
 LSAKFAQNFDAQHYEASLTGCAANNPSITLPAQDPVYVAGAGTGTGHTTTQAAVD
 AAIKRTKROYIAVMPGYGTIVYPAPGCTITLTGTEKPIDVKIGISLUGGSPA
 DMRHDNPKRGATMPGKPAWYMYDSCQSRSDIGVLSAVENSGNNGLOLTLTENT
 LGSIDVADNHPVALRTDGDQOINNVNITIGRQNTPEYVNSGVNRLETRNPRLTV
 NSYIDGVDIVSGCAVVDNFEFRVNSRTQEAAYVAPATLSNITYGFLAVNSREN
 ATGCGVAOLGRLSDVDANTNGOVIRDSAINEGFMARPMADAVVLSNRPAGNTGSDV
 DNDTEORNLNDTNRMEXNNRGVSKYVAAKK"
 complement(1507..1535)
 /note="factor Sigma70; predicted +1 start at 806574"
 complement(1582..2058)
 /gene="ydhb"
 /note="b0773"
 complement(1582..2058)
 /function="orf; unknown"
 /note="f158; 99 pct identical to YBH_ECOLI SW: P12994"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC73860.1"
 /db_xref="GI:1786990"
 /translation="MKLISNDLRDGDKLPRIHHEVNGMGVGDGNTSPLHAMDVPAGTK
 SEVVTGDPDAPGTSGMMHWVYVNLPAQTRVLPQGFSGSLVAMPDGVLOTREDFEKTG
 YDGAAPKGETHRHYETVHALDIERIDVDEGASGAMVGFNVHFLASASTAMFS"
 complement(2108..2124)
 /note="central position to predicted promoter:85.5"
 /bound_moiety="Rhas predicted site"
 complement(2117..3406)
 /gene="b10A"
 /note="b0774"
 complement(2117..3406)
 /gene="b10A"
 /EC_number="2.6.1.62"
 /function="enzyme; Biosynthesis of cofactors, carriers:
 biotin"
 /note="f429; 100 pct identical to B10A_ECOLI SW: P12995"
 /codon_start=1
 /transl_table=11
 /product="7,8-diaminopelargonic acid synthetase"
 /protein_id="AAC73861.1"
 /db_xref="GI:1786991"
 /translation="MTTDIAFDORHWHPTMTSPUPYVVASGCECILSDGR
 LVDGSSWMAIHGYNHPOLNAAKMSQDANSHVFGSITAPALECRKLVAMRPOP
 LECVFLADSGVAEVAVKMALQIWAQGEARORPLFRNGYHGDTEGAMSVCDPNS
 MHSIMKGYLPENLEFAPQSRMDGEMDERDVGFRILMAARHETIAVILPIYOGAG
 MSHMYHPMLKRIKICDREGILLIADETATGFGTKLFACEHETIADILCLGKAL
 TGGITLTATLTREVAETISNGEAGCFMGPTGPNPLACAAVASLALLESGBWOO
 QVADIEVOLREOLAPADAEMVADVRLGATGVETTPVMAALOKRFFVEGWIIR
 FGKLIYMPYIILPOOLORLTAAVNRVODTEFFCQ"
 complement(2193..2221)
 /gene="b10A"
 /note="factor Sigma70; predicted +1 start at 807260"
 3411..3450
 /note="central position to biob promoter: -20"
 /bound_moiety="Biob documented site"
 3411..3450
 /note="central position to predicted promoter: -20; genetic
 evidence for the site"
 /bound_moiety="Biob predicted site"
 complement(3411..3450)
 /note="central position to predicted promoter:50"
 /bound_moiety="Biob predicted site"
 complement(3411..3450)
 /note="central position to bioA promoter:50"
 /bound_moiety="Biob documented site"
 3413..3441
 /note="factor Sigma70; promoter biob: documented +1 at
 808525"
 complement(3447..3473)
 /note="factor Sigma70; promoter bioA; documented +1 at

808515"
 3493..4533
 /gene="b10B"
 /note="b0775"
 3493..4533
 /gene="b10B"
 /EC_number="2.8.1.-"
 /function="enzyme; Biosynthesis of cofactors, carriers:
 biotin"
 /note="o346; 99 pct identical to B10B_ECOLI SW: P12996"
 /codon_start=1
 /transl_table=11
 /product="biotin synthesis, sulfur insertion?"
 /protein_id="AAC73862.1"
 /db_xref="GI:1786992"
 /translation="MAHRPMTLSQVTELEKPLDLLFEAQVHRQHPDPROYST
 LLSIKTGACPEDEKCYQPSQSRKYGLEARLMEVEQVLESARAKAGSTRFCGAAAM
 KNHREDMPYLEQWVGKAMGLEACAGTLCGLTSGOARLANGLDYLNLPTSPER
 YGNITRTYGERLDLEKVDAGIKVCSGIVGICPTVDRAGLLIOLANLPPEPS
 VPINMLYKACSTPLADDVADPFTITAVARIMPTSYRLSAGREONREQTQAMC
 PMAGANSITFGCKLITTPNEBEDKDLQLFRLGLINPOQTVAVLAGDNDOQRLBOALMT
 PDIDBYNAAL"
 4530..5684
 /gene="b10F"
 /note="b0776"
 4530..5684
 /gene="b10F"
 /EC_number="2.3.1.47"
 /function="enzyme; Biosynthesis of cofactors, carriers:
 biotin"
 /note="o384; 100 pct identical to B10F_ECOLI SW: P12998"
 /codon_start=1
 /transl_table=11
 /product="8-amino-7-oxononanoate synthase"
 /protein_id="AAC73863.1"
 /db_xref="GI:1786993"
 /translation="MSWQEKINMALDARRAADLRRRYPVAGAGRLVADDOYINF
 SSNDYGLSHHPQIRAMOGGADPFGSGSGHSYXSVYHOLEBELAEMIGYSRA
 ILPTSGAANDAVTAAMAKEDRIADRRLSHASILEKASISPSQLRFANDVTHLAR
 LLASPCPGQGVVTEGVSMDCGSAPLAEIQOVTQONGVLMVDAGHGVIGEGDGR
 SCWLQKVPPELLVTFGKFGVSAALVCSSTVADYLLQFARHLIYSTMPAQOAL

Query Match 100.0%; Score 25; DB 1; Length 11022;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtgcgtgacatcgtctgttgaagc 25
 |||||||||||||||||||||
 Db 3549 GTTGCTGATCTGCTGTTGAAGCG 3573

RESULT 14
 AE005258
 LOCUS
 DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
 of 155.
 ACCESSION AE005258 AE005174
 VERSION AE005258.1 GI:12513751
 KEYWORDS
 ORGANISM
 Escherichia coli O157:H7 EDL933.
 Escherichia coli O157:H7 EDL933
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia
 1 (bases 1 to 13501)
 Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
 Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
 Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
 Grobbeck,E.J., Davis,N.W., Ijm,A., Dimalanta,E., Potamousis,K.,
 Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
 Welch,R.A. and Blattner,F.R.
 Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
 Nature 409 (6819), 529-533 (2001)

TITLE
 JOURNAL MEDLINE
 21074935

```

PUBMED 11206551
REFERENCE 2 (bases 1 to 13501)
AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rosen,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grobeck,E.J., Davis,N.W., Llm,A., Dimmalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
source 1. 13501
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohemorrhagic"
misc_feature <1..7576
/note="O-island #36; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655; Cryptic prophage
CP-933K; Includes one copy of the 13 bp direct repeat that
flanks the prophage"
66..665
/gene="1omk"
/note="20981"
66..665
/gene="1omk"
/function="putative membrane; Other or unknown (Phage or
prophage Related)"
/note="Residues 1 to 199 of 199 are 72.86 pct identical to
residues 1 to 199 of 199 from Genpept 118 :
g17532789|gb|AA63231.1|AF151091_2 (AF151091) Lom
|prophage P-E1bA]"
/codon_start=1
/transl_table=1
/product="putative outer membrane protein of prophage
CP-933K"
/protein_id="AAG55137.1"
/db_xref="GI:12513752"
/translation="MKVCAIILSAICLAVGVPANASEHSTLSAGVILHASTDAG
SDLLNGINVKRYEFTDILGLITSFYANMEDQKTRYSOTRMHEDYVRNRWFSVAG
PSRVNEMFSAAGAVASRVSTSGDYFRTDNKRKTHDVLTSDDARYSMTSLAW
GAGVQFNFTESAVDAVAYEXSGSGDMRTDGIYGVGYK"
719..2041
/gene="20982"
719..2041
/gene="20982"
/function="putative structure; Structural component (Phage
or Prophage Related)"
/note="Residues 164 to 440 of 440 are 68.79 pct identical
to residues 381 to 645 of 645 from Genpept 118 :
g14585436|gb|AAD25464.1|AF125520.59 (AF125520) putative
tail fiber protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=1
/product="putative tail component of prophage CP-933K"
/protein_id="AAG55138.1"
/db_xref="GI:12513753"
/translation="MGMAAVOISGVCLKDGAQKPIQNTQLKARNSTXVVNTVASE
NPDEAGRYSMDEVGYOYSVTLVVGFPSPSAGTISYEDSOPGTLNDFLCAMVEDAR
PEARLRFQWYEAARHAEBAKKNAGAEISARNAGISASAKAEASANADTSAMEDASE
SAROAESAAGAKSEASSEASSASAEAKASESLOSATDAELSKTSAAGNAARA
TTSERKASAESENOASOSRIAEADAVNRILPYVGPGRKGEPAQPGCPGCKDQGE
RGDTGPACATCEBPGSGDTGPACPGKPGKGRGERGEGTLGNAGPGCPKDDTCAQPA
GPGCPKETGAAGVGVATGPGKPGDGETGQIRIRLGPMTITFNSTGWPPTDGAALI
TGTLFDPKDQATQYGMFOHLQVHFGPGDQVQKGLDEVDGSDTRTGE"
2004..2312
/gene="20984"
2004..2312
/gene="20984"
/function="orf; Other or unknown (Phage or Prophage
Related)"
gene
CDS
/note="Residues 1 to 102 of 102 are 98.03 pct identical to
residues 28 to 129 of 129 from Genpept 118 :
g14585437|gb|AAD25465.1|AF125520.60 (AF125520)
hypothetical protein [Bacteriophage 933W]"
/transl_table=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55139.1"
/db_xref="GI:12513754"
/translation="MKMAVIOALEMDNMILKKLQSLGCGKHCDCNGRSRTAQLR
LGAADILESDENGIIPEDQRTVITQVYLLADKKQIQGVVRPLQILRADGTWENIGCMK"
2489..3469
/gene="20985"
2489..3469
/gene="20985"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 13 to 326 of 326 are 55.73 pct identical
to residues 18 to 331 of 336 from Genpept 118 :
g16960367|gb|AA633527.1| (AF170176) hypothetical protein
predicted by glimmer [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55140.1"
/db_xref="GI:12513755"
/translation="MLSPITRTTFHNSVINIVOSPCQVTSFPAKEKELVDEKPIFL
QWEENPERYKKDEVIYNTKQHPYLDVNTNARIESDMIGIFVDGPFVSQKAFS
KLERDENPVIYREDVDPSMYDRKLSIDYHDIICEORLRTEDKDEYLLNLEKELR
EISKADDSLSYAKRRHNAWDFPFLALAKAGEIPCTQYKKNHGISFGGCGLYLD
MMILUNGKLSITVADGISMHYDRRDSVNIENSAIIVRSNHPALLGLGSPMHSKVD
AHPYDGLCKGYKAKTFNFTPLNKNHCFEFNPNITMNTSQTCSW"
3330..4522
/gene="20986"
3330..4522
/gene="20986"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/transl_table=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55141.1"
/db_xref="GI:12513756"
/translation="MKIPSLQSNFNSPAGSAPAPARNAENAYADVLDIGKRIPL
SAADISNVEESVIRAVHDSRRLIDQHTVDMIGNTVLALSSQTFRDVAVSYGHNEK
VHIGCTIKYRNEYELNEESSVKIDDIOSLTCNLSLYEDVGOEIPICIEAGRNDDEPY
VESVAPDIDSYEMPSWOGELIHETIIHNTGSSDPSGSNIIEIGFTEILARRVQDELG
WSPDPRGYAEEPEREHLRLRLNLRLQAMHEHENERAFPERLGTITSDREASDPFT
EYSVSNIGYGFILQCHDFPGLAINNLDQANOIOLYHGAPVITFTGVDKHNQO"
5350..6231
/gene="20989"
5350..6231
/gene="20989"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55142.1"
/db_xref="GI:12513757"
/translation="MLSPSVNLGCSNLSIRNLTSPPNRVLSYVRDAVHSDNGAOY
KGNKRTYRVATDNKFCYTRRESHSCPTNLNLRGCPKGETISRKTEVNLNAPSAAH
ENGIVHSNRPDLPPVDVAPPELPSVDYRNLSPHVGIGKGNAAVYEDAEADATKLMK
FTTSQSNIEVTSERCFNQYGAAGAEKTYGNNGDIIGIRMDKINGESLNLISLPAQ
AEHAITYDMEDRLQCKGILFVDTETNVLVYBRAKNEFNIDISSYVSDRSSESOIMO
SYHGKODLISVYLSKI"
6462..7160
/gene="20990"
6462..7160
/gene="20990"

```



```

/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="Unknown protein encoded by prophage CP-933K"
/protein_id="AG55143.1"
/db_xref="GI:12513758"
/translation="MRPSTNIVLHSSSSSSSPDIDSLKASSVQIKNNPOLRE
QGTDDHNYOOIEALDKIGSTETGRVLLNIESISLKESTVYIHNSSRLGVMAHD
IDAEHRTGSDFCNLMNAVEPCGGISVDPHATVFEHLHFNHNGELKVES
SRAESQKSYSLPLEARVQIGAFSEEVLSSEKFEHIGMPRTSYPSXSLIHDDNT
VSLGPOOVRLLPIL."
/complement(7668.8144)
/gene="ybH8"
/note="20992"
/complement(7668.8144)
/gene="ybH8"
/function="orf; Unknown function"
/note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 strain
MG1655: B0773"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AG55144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLRDGLKPLRHVFNCGMGVGDGNISPHLAMDVPACTK
SPVTCYDPAIDAPGSGMHWVYNLPADTRVLPQGGSLVAMPDVLQTRDPFGKTS
YGCAPPKGEIHRITFTVHALDERIDVDEGASGAWGNVHFSLASITAMFS"
/complement(8203.9492)
/gene="bioA"
/note="20993"
/complement(8203.9492)
CDS
Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 gtctgcatctgtctgttgaagc 25
Db 9635 ctgcctgcatctgtctgttgaagc 9635

RESULT 15
AP002553 297816 bp DNA BCT 07-MAR-2001
LOCUS Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION AP002553
VERSION AP002553.1 GI:13360211
KEYWORDS
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub-strain:RIMD 0509952)
DNA.
REFERENCE Escherichia coli O157:H7
AUTHORS Bacteriophage: Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (sites)
Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C. H., Kimura, S.,
Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T.,
Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
Sasakawa, C., and Shinagawa, H.
Complete nucleotide sequence of the prophage VP2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
2 (sites)
Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M.,
Hayashi, T.,
Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and
Hayashi, T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655

JOURNAL Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
MEDLINE 20557356
REFERENCE 3 (sites)
AUTHORS Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S.,
Yutsudo, C. H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and
Shinagawa, H.
Complete nucleotide sequence of the prophage VP1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
4 (sites)
AUTHORS Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T.,
Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C.,
Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and
Shinagawa, H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
21156231
5 (bases 1 to 297816)
AUTHORS Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and
Hayashi, T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center: 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail: ken@gen-info.osaka-u.ac.jp,
URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,
Fax: 81-6-6879-2047)
genome project.
COMMENT
FEATURES
LOCATION/Qualifiers
1..297816
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
79..1245
/gene="ECS0753"
79..1245
/gene="ECS0754"
/note="similar to SUCC_ECOLI g111786948 percent identity
100 in 388 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="succinyl-CoA synthetase beta subunit"
/protein_id="BAB34176.1"
/db_xref="GI:13360212"
/translation="MNHEHYAKQLFARYGTPAVGYACTTPPEAEASKIGAGPWV
VKCVHMGGRKAGGVKRVNSKEDIRAFENWGLKRLVTVOTDANGQPVNOLIVEAT
DIARELYGAVVDRSSRRVFMASTEGVELEKVAEETPHLIRKVALDPLTGPYOG
RELAFKLTIEGKLVQOFTKIIPMGATITFLEBDLALIEINPLVITKQDGLCLQKGA
DGNALFROPDLREMRDQSDPEPRAQAOEMLVNVALDNGICWVNGAGLGMMDIV
KLHGGEPANFLDVGAGATKERVTAFAFIIISDDKRVAVINIRGCIYRCOLLADGII
AVAEVGVVPPVVRLEGNMELGAKKIADSLNIIIAKGLTDAAGVVAVEK"
1245..2114
/gene="ECS0754"
1245..2114
/gene="ECS0754"
/note="similar to SUCD_ECOLI g111786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/translation="MSTLIDKNKRVITGCGTSGGGRPHSFOALATGKMGVYTPGK
GTTHLGIPVNTVREAVAAAGATASVIVPAPFCQDSILEADAGIKLITTEGIPPT
LDMLTIVKVLIDEAGVIRMGICNPQVITPGECKIGIQGHIRPGKVGIVRSCTLYE
AVKOTTDYGRGOSTCVGIGDPIPGSNFIDILMFEXKDPQTEAIVMIGETGSAEEBA

```

```

gene
CDS
    2781..3686
    /gene="ECs0755"
    /note="probable transcriptional regulator, similar to
    transcriptional activators e.g. glycine cleavage system
    transcription activator (gcv operon activator) -
    Escherichia coli g114170431sp|p32064|GCV_A_ECOLI percent
    identity 31 in 300 aa"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="putative transcription regulator"
    /protein_id="BAB34178.1"
    /db_xref="GI:13360214"
    /translation="MRCKIPKTELLVFEVVAHREHSTVRAEELATLQSAVPROVAL
    EEPFLPLFSSKKRIFLNDGKYTLGIVKTTAKLEBDRTTITWOPPYQVITLAVN
    PFSTHWLIPMLHEFTKLHPDIIVNHSLSNNGDFLNREYDAVIRENFCAPAEVY
    LFEELIPLVCSGLSAMSOKLSVAELTTEPLLHQSTRITGMEWEMFALGVSSPLVN
    NGPREDLSMLIAVRSNIGVALLPFAIQLDLDSCDMVIPCQVPIRTGKRFTWQFE
    EKSDSPHLOQFREWILAKSVYPOEM"
    complement(3720..4322)
    /gene="ECs0756"
    complement(3720..4322)
    /note="ECs0756"
    /note="probable cob(I)alamin adenosyltransferase, similar
    to cob(I)alamin adenosyltransferases (cobI)
    adenosyltransferases) e.g. [Escherichia coli]
    g11151481sp|p130401|BTUK_ECOLI percent identity 67 in 200
    aa"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="putative cob(I)alamin adenosyltransferase"
    /protein_id="BAB34179.1"
    /db_xref="GI:13360215"
    /translation="MEARISTEHRVQROQKLEQVDTRVAALIEKKGLIVFGNGKG
    KSTAAFGTVTRAVGHGTVGAQYTKGQMDGYNLQPLQGVFPHMGTFWQTKOR
    QADIDAAKEVSESKRMALADKRDYLVLDLITMLAYHLDTEEVIASLONRPAQDSY
    IVTGSGHSQLKMAQDVSEIRPVKHAFFDNGICQAPGIDW"
    complement(4332..5984)
    /gene="ECs0757"
    complement(4332..5984)
    /note="ECs0757"
    /note="probable fumarate hydratase, similar to fumarate
    hydratases e.g. fumarate hydratase class I, aerobic
    (fumarase) - Escherichia coli
    g11205981sp|p00923|FUMA_ECOLI percent identity 68 in 545
    aa"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="putative fumarate hydratase"
    /protein_id="BAB34180.1"
    /db_xref="GI:13360216"
    /translation="MSKPFITWQEPFLONKQDEYTLISDQHTTTELDGEVRIKAE
    ALTLARAFTEASFPLRSALHDOVASILNDQSSANDKYALDOLLNRAEVSAGVLP
    NCDDGTATVIAVSGOOITWGGNDPAEALSIGYSTFOENLRFSONAPLDMYEVNTO
    TNPDAQIDISAVAGDEYHFLCVNKGGSANKAALYQETKSLQPEKLTAFLEKVL
    GTACPFYHIAFVVGVSADQTLKIAKLASTKYVDNLPTSGNBOGAFRDIELEKVL
    EASOFGIGAOFGKGFVADIRVLRPHGSGCPIAMLSGNSADRNRKAKNKGIYL
    EKLEHNGQYIPASIRENNHGHQDLNRPIDRMODIARLPVGTSTVSLSGPIVAVR
    DIAHAKIKARLDSEPMPEYIKHHIYVYAGAPKPKENACSGLEPTTGGRRDQITPE
    QAGGSLVMSKNSQOYTDACHRHGGEFNLSGTGAALALAQETVSLRCLLEYBELG
    MEAVMMLEVENLPAFLVDDKGNFFSOEQDHRACSPACH"
    complement(6092..7372)
    /gene="ECs0758"
    complement(6092..7372)
    /note="ECs0758"
    /note="probable transport protein, similar to
    glutamate/aspartate transport proteins (proton glutamate

```

```

symport proteins) e.g. [Bacillus stearothermophilus]
g11214671sp|p24943|GLT_BACSF percent identity 38 in 416
aa, also similar to C4-dicarboxylate transpor"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="putative transport protein"
    /protein_id="BAB34181.1"
    /db_xref="GI:13360217"
    /translation="MKKISLTMTMLIALVGMIGVILNMTASPEAKLYAOEISIFT
    TTEFLRKIMTIAPIVYSTLIVGVIJAKGDAKALGIESKTLFELICASLSIALGLITV
    NPEPSTGTFNFAVHGAETTVGVAAPETLKVISHAPPSIVDAMAHBEILOIYVESI
    PIAGSITATGEGSAIVAHDLDAHAKLKTQYVNLPAPIYPAASIALIAERGLAV
    VSGITPMEFTFTMLLWVLIGLIVVGPICIRUPLALSPLALATTTSSSEAPF
    GTLEKLEQGVSPKIASFVLPITGSEFNLVGSVAFFATVIAQCNIHLSTGEQITM
    LILMLITSGMAGVPRASMVVIAATLNDPNIPEAGLILMGVDPFLDGRSATVMSN
    AMGAAMVSRMEGHEFCEGCRKALKPESNALP"
    complement(7533..7850)
    /gene="ECs0759"
    complement(7533..7850)
    /note="ECs0759"
    /note="unknown"
    /codon_start=1
    /transl_table=11
    /evidence=not_experimental
    /product="hypothetical protein"
    /protein_id="BAB34182.1"

Query Match      100.0%; Score 25; DB 2; Length 297816;
Best Local Similarity 100.0%; Pred. No. 0.00097;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gttgctgacatcgtctgttaagcg 25
    |||||
Db 92367 GTTGCTGATCTGCTGTTGAAGCG 92391

RESULT 16
E00893 LOCUS 1121 bp DNA PAT 29-SEP-1997
E00893 DEFINITION Genomic DNA encoding biotin Synthetase.
E00893 ACCESSION E00893
E00893 VERSION 1
E00893 KEYWORDS JP 1986149091-A/1.
E00893 SOURCE Escherichia coli.
E00893 ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 1121)
AUTHORS Hirono,Y., Kojima,T. and Kimura,H.
TITLE DUPLEX DNA TO CODE BIOTIN SYNTHASE. BACTERIUM CONTAINING SAME AND
PRODUCTION OF BIOTIN
JOURNAL Patent: JP 1986149091-A 1 07-JUL-1986;
COMMENT NIPPON SODA CO. LTD
OS Escherichia coli
PN JP 1986149091-A/1
PD 07-JUL-1986
PF 24-DEC-1984 JP 1984272605
PI HIRONO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
C12N15/00, C12N1/20, C12P13/18, (C12N1/20, C12P13/18), (C12P13/18, PC
C12P13/19);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: strain=Escherichia coli Ns101;
CC feature is identified by experimental;
FH Key Location/Qualifiers
FT CDS 42..1079
FT /product="biotin synthetase".
FT location/Qualifiers
1..1121

```

```

/organism="Escherichia coli"
/db_xref="taxon:562"
BASE COUNT      289 a      296 c      325 g      211 t
ORIGIN

Query Match      96.0%; Score 24; DB 10; Length 1121;
Best Local Similarity 100.0%; Pred. No. 0.0055;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1      gttgcgtgacatcgtctgttgaagc 24
          |||
Db      98      GTTCTGATGATCTGCTGTTGAAGC 121

RESULT 17
LOCUS      AE004192      12891 bp      DNA      BCT      31-JUL-2000
DEFINITION      Vibrio cholerae chromosome I, section 100 of 251 of the complete
ACCESSION      AE004192 AE003852
VERSION      AE004192.1 GI:9655581
KEYWORDS
SOURCE      Vibrio cholerae.
ORGANISM      Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE      1 (bases 1 to 12891)
AUTHORS      Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
            Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.,
            Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
            Ermolaeva, M.D., Yamatchevan, J., Bass, S., Qin, H., Dragol, I.,
            Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D.,
            Nierman, W.C. and White, O.
            DNA sequence of both chromosomes of the cholera pathogen Vibrio
            cholerae
            Nature 406 (6795), 477-483 (2000)
            20406833
TITLE      2 (bases 1 to 12891)
JOURNAL
MEDLINE
REFERENCE
AUTHORS      Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
            Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.,
            Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
            Ermolaeva, M.D., Yamatchevan, J., Bass, S., Qin, H., Dragol, I.,
            Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D.,
            Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,
            Mekalanos, J.J., Venter, J.C. and Fraser, C.M.
            Direct Submission
            Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
            Location/Qualifiers
            1..12891
            /organism="Vibrio cholerae"
            /strain="N16961"
            /serotype="O1"
            /db_xref="taxon:666"
            /chromosome="1"
            /note="biotype: El Tor"
            complement(84..1370)
            /gene="VC1111"
            complement(84..1370)
            /gene="VC1111"
            /note="similar to GB:J04423 SP:P12995 PID:457106
            PID:490217 GB:U00096; identified by sequence similarity;
            putative"
            /codon_start=1
            /transl_table=11
            /product="adenosylmethionine-8-amino-7-oxononanoate
            aminotransferase"
            /protein_id="AAF94270.1"
            /db_xref="GI:9655582"
            /translation="MMDLADRCIHMPYSTLTPLTCYVSADGVMITLEDGRSL
            VDGSSWMAITHGNIHKLNQAHQIOMSHIMGGTLTHQPALELCOKLKLAIPNML
            COVLADSGSAVAEVSILKALQYMSKGPAPKPLTLRHHGYHGTFFAMASVTDPDSM
            HSLVXGFLPEHIFANSPEGFFDAMDERDIADPFRHKLTENHNOIAAVTLEPIYOGACG

```

```

gene
CDS
1497..2549
/gene="VC1112"
1497..2549
/gene="VC1112"
/note="similar to GB:J04423 SP:P12996 PID:145425
PID:490219 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="biotin synthase"
/protein_id="AAF94271.1"
/db_xref="GI:9655583"
/translation="MEVHHMTVAEVRKALDLKPRMDLFEAQGVHRLHHPHNVOST
LLSIKTAGCPEDCKYCPOSAHYPTDVNKRLEMEVERVLDAOKAKNGSTRPCGAM
KNPERDMPILKEMIRGVKMDGLCTCTGCGILPPDOALQAGLDVYNNHIDPSPEF
YGNITITRTYDRLDITLSHRDAGMKICSGGIIOMGSTNDRLAGLVELANLTPRES
VPIMLVKVKGTPLEQYDVEPEFVFLIARIMPKSAVRUSAGREKNEQWALC
FMAGANSIFYGCKLITTPNPAEDSDMLFKLGINRQVAKQPEITENELDRYVER
VAARPTASDLFYDAAL"
2533..3687
/gene="VC1113"
2533..3687
/note="similar to GB:J04423 SP:P12998 PID:145426
PID:490225 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="8-amino-7-oxononanoate synthase"
/protein_id="AAF94272.1"
/db_xref="GI:9655584"
/translation="MTLRFKSRIRIOLNREHSHLTPRLQLDPSAGGELLIOODRYL
NESSNDYLGANDAEIVQAMOTGLARFGAGSGASPMWTGFSAAHEHLKTEL
RAVFESSGFSANQALFTLEKEEDLLQDLNHNASLMEAKMSLRATKRRKNDTEL
ROLLHEGNSLVTEGYFSMDGDCAPLAQIATLVKORAMLMVDAAHIGVLEDGAS
COAGIHPELIVTEGKAFGLAGAAVLCDAAHVGDTLQFARHHVYSTAMPPOAHMLT
HAAEMIOSQSWRREBLAELACPEDCRNIPGVATQTPKPMILGSSDSLSASHEL
KQOQIWSAIPRPVPGSARLRITTYAHTQAOIRLSBQLKHYMEKDT"
3549..4487
/gene="VC1114"
3549..4487
/note="similar to GB:J04423 SP:P12999 PID:145427
PID:490221 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="biotin synthase protein bioc"
/protein_id="AAF94273.1"
/db_xref="GI:9655585"
/translation="MGECSTTHRACROGACATYHDNGGAYPASDTPILGAVKCDGEG
HMSMTATELCVELKRCASIAOAFGKAAHFDQAAARPOGGLLOKMSCELGAVL
DLGGCTGYFSLRLBERGAQYVCADISAMLEQAKORCGDGMSTQLADAEQLPASAC
FDWVFSSLAQWCEADLSLSEIRRVLYKPRGQAFLLSTLDGSLFELEQARVSDHNRH
INQFISINQVYKIALAQAGCSQHNHDLAATVWAEAFMLRDLKIGANHVSGRSTGL
ISRTTLAKVELAVQSPFRNQGGLVPATGYVCLGVYIHR"
4463..5179
/gene="VC1115"
4463..5179
/gene="VC1115"
/note="similar to GB:S68059 SP:P13000 PID:490223
PID:145428 PID:460901; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase"
/protein_id="AAF94274.1"
/db_xref="GI:9655586"
/translation="MFRGYSPMRNALFIAGTDVDGKTVASKAIIQALATHNATIGY

```

KPAAGSDKTEEGYRNSDAHLAKAATVMPEDVNPYALVLPSPHIAKHENVTID
 YALLSNKLSKRONAELVVEAGWVRPTSDSCSLSTWKOERLPVTLVVGTKLGL
 SHALITAEAIRADGILELVGWIARINPGETHYAIIIEHLBGRIGLTPKIGELPYMPKAK
 ROELGFIOLDHLEPDTVA*
 complement(5261..5470)
 /gene="VC1116"
 complement(5261..5470)
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="AAP4275.1"
 /db_xref="GI:965587"
 /translation="MSRRNLITNKGHMAEASVYOMADLLCCHGISPOEARTOLCL
 EAKQTLQOOIVDTQSHLMGISAER"
 /gene="VC1117"
 5697..6560
 /gene="VC1117"
 /note="similar to GB:M58470 SP:P23894 PID:146412 GB:U00096
 PID:1736470; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="heat shock protein Htpx"
 /protein_id="AAP4276.1"
 /db_xref="GI:965588"
 /translation="MKRIILFLATINLAVLVSLVNIYVAVTGMQPGSLGLIWMVA
 VEGFGAFISLMSKSMALRSVGVVDTPENEHMLLETVRQAQACIGMPTVAI
 YDADPMNAFATGAKRDSLSVAVSTGLNDRDAEAVLAHVESHANGDMPTLMO
 GVNTFVFLSRFLAVASRDESGESNMVYGVSVNLELVEFGFASFTIMVSR
 HREPHADGAAGLQVYKHKMIAALERLKMGEHLESGMAFGITGRKSLSLMTHTP
 LEKITALRNM"
 /gene="VC1118"
 6770..7198
 /gene="VC1118"
 /note="similar to PID:1653608; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="transcriptional regulator, putative"
 /protein_id="AAP4277.1"
 /db_xref="GI:965589"
 /translation="MSIMDVQVNAQFOQLDKSQLHRLITETHPDVVFEDAAHRTGCF
 DALVOYFLNTOYNTCTPTTHOYAVNEGAFVMTMLRHPKLAKEQVDVGVSHL
 HFAGKYVHRDIDYDMGEMLYEQLPVLCGVTRAIKRLGQ"
 /gene="VC1119"
 7195..7920
 /gene="VC1119"
 /note="similar to SP:P25970; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="oxidoreductase, short-chain
 dehydrogenase/reductase family"
 /protein_id="AAP4278.1"
 /db_xref="GI:965590"
 /translation="MTAVPTTGATSGTGKQLADYAKQGVYACGNQVDSLSHQ
 YANIFPLAFDVTDPHTKALADQLCOPELMLNAGCEITIDGKMPVTLMAVFNLN
 VLGVAAYIEGILQPHLSGHRVATVGSIASELAPRAVYASRAAVYIARTLQDLNR
 PLGEVETIIPGPAVPLTDNPFAMPMITVERAAOEIKAGLAGVSOLYFPKRFPM
 LIRLIGALPYMGRVLRLLKA"
 /gene="VC1120"
 7931..9211
 /gene="VC1120"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=1

Query Match

80.0%; Score 20; DB 1; Length 12891;

Best Local Similarity 100.0%; Pred. No. 0.65;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 tgcgtcgtcgttgaagcg 25
 |||||
 Db 1558 TGCATCTGCTTTGAAGCG 1577
 RESULT 18
 S64534/c 383 bp DNA INV 27-APR-2000
 LOCUS
 DEFINITION
 (dec-1) gene, dec-1-Fc4 allele, partial cds.
 ACCESSION
 S64534.1 GI:407575
 VERSION
 KEYWORDS
 SOURCE
 ORGNAMISM
 fruit fly.
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE
 AUTHORS
 1 (bases 1 to 383)
 TITL
 Evolution of the dec-1 eggshell locus in Drosophila. II.
 Intrasppecific DNA sequence analysis reveals length mutations in a
 repetitive region in D. melanogaster
 JOURNAL
 J. Mol. Evol. 36 (6), 536-544 (1993)
 MEDLINE
 93353525
 PUBMED
 8350348
 REMARK
 Genbank staff at the National Library of Medicine created this
 entry [NCBI gblseq 136269] from the original journal article.
 This sequence comes from Fig. 4.
 Map location: X 7C3-4.
 FEATURES
 source
 1..383
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 <1..>383
 /gene="dec-1"
 /allele="Fc4"
 <1..>383
 /gene="dec-1"
 /note="This sequence comes from Fig. 5; follicle cell
 protein for proper eggshell assembly; conceptual
 translation presented here differs from translation in
 publication"
 /codon_start=1
 /product="defective chorion-1 eggshell protein"
 /protein_id="AAB27805.2"
 /db_xref="GI:7548373"
 /translation="ONPMIMQORWSEEQAKIQONQOIOONPMIMQORWSEEQAKI
 OONQOIOONPMIMQORWSEEQAKIQONQOIOONPMIMQORWSEEQAKI
 QQAQQAQOQTPMMPMOMQORWTEPP"
 BASE COUNT 143 a 90 c 102 g 48 t
 ORIGIN
 Query Match 68.0%; Score 17; DB 96; Length 383;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 ttcgtcgtcgtcgtt 18
 |||||
 Db 159 TTGCTGATCTGCTGTT 143
 RESULT 19
 S64532/c 461 bp DNA INV 27-APR-2000
 LOCUS
 DEFINITION
 Drosophila melanogaster defective chorion-1 eggshell protein
 (dec-1) gene, dec-1-Fc3 allele, partial cds.
 ACCESSION
 S64532
 VERSION
 S64532.1 GI:407573

RESULT 24
 DSDC1TR6/c 774 bp DNA INV 25-SEP-1996
 LOCUS D.sechellia DNA from repeated region of dec-1 locus.
 DEFINITION X96931.1 GI:1262763
 VERSION X96931
 KEYWORDS dec-1 locus; tandem repeat.
 SOURCE Drosophila sechellia.
 ORGANISM Drosophila sechellia.
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 774)
 Escher, S.A. and Lambertsson, A.
 Evolution of the dec-1 eggshell locus in Drosophila. III. Sequence
 comparisons of the simulans complex repeated domain reveal
 non-concerted evolution
 JOURNAL Hereditas 124 (2), 111-120 (1996)
 MEDLINE 96376325
 REFERENCE 2 (bases 1 to 774)
 AUTHORS Escher, S.A.
 TITLE Direct Submission
 JOURNAL Submitted (22-MAR-1996) S.A. Escher, Umea University, Department of
 Genetics, S-901 87 Umea, SWEDEN
 COMMENT Related sequences: M35889, M35688 and X07697.
 FEATURES
 source
 1..774
 /organism="Drosophila sechellia"
 /db_xref="taxon:7238"
 /chromosome="X"
 /map="7C"
 repeat_region 1..774
 /rpt_type="TANDEM"
 gene 1..774
 /gene="dec-1"
 exon <1..>774
 /gene="dec-1"
 /number=2
 BASE COUNT 293 a 178 c 213 g 90 t
 ORIGIN

Query Match 68.0%; Score 17; DB 6; Length 774;
 Best local Similarity 100.0%; Pred. No. 35;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ttgctgacatcgtcgtt 18
 ||||||||||||||||
 Db 81 TTGCTGATCTGCTGTT 65

RESULT 25
 DSDC1TR2/c 852 bp DNA INV 25-SEP-1996
 LOCUS D.simulans DNA from repeated region of dec-1 locus, strain
 DEFINITION YUGOSLAVIA.
 VERSION X96927
 KEYWORDS dec-1 locus; tandem repeat.
 SOURCE Drosophila simulans.
 ORGANISM Drosophila simulans.
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 852)
 Escher, S.A. and Lambertsson, A.
 Evolution of the dec-1 eggshell locus in Drosophila. III. Sequence
 comparisons of the simulans complex repeated domain reveal
 non-concerted evolution
 JOURNAL Hereditas 124 (2), 111-120 (1996)
 MEDLINE 96376325
 REFERENCE 2 (bases 1 to 852)

AUTHORS Escher, S.A.
 TITLE Direct Submission
 JOURNAL Submitted (22-MAR-1996) S.A. Escher, Umea University, Department of
 Genetics, S-901 87 Umea, SWEDEN
 COMMENT Related sequences: M35889, M35688 and X07697.
 FEATURES
 source
 1..852
 /organism="Drosophila simulans"
 /strain="Yugoslavia"
 /db_xref="taxon:7240"
 /chromosome="X"
 /map="7C"
 repeat_region 1..852
 /rpt_type="TANDEM"
 gene 1..852
 /gene="dec-1"
 exon <1..>852
 /gene="dec-1"
 /number=2
 BASE COUNT 324 a 195 c 232 g 101 t
 ORIGIN

Query Match 68.0%; Score 17; DB 6; Length 852;
 Best local Similarity 100.0%; Pred. No. 35;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ttgctgacatcgtcgtt 18
 ||||||||||||||||
 Db 237 TTGCTGATCTGCTGTT 221

RESULT 26
 DSDC1TR3/c 852 bp DNA INV 25-SEP-1996
 LOCUS D.simulans DNA from repeated region of dec-1 locus, strain
 DEFINITION BARCELONA.
 VERSION X96928
 KEYWORDS dec-1 locus; tandem repeat.
 SOURCE Drosophila simulans.
 ORGANISM Drosophila simulans.
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 852)
 Escher, S.A. and Lambertsson, A.
 Evolution of the dec-1 eggshell locus in Drosophila. III. Sequence
 comparisons of the simulans complex repeated domain reveal
 non-concerted evolution
 JOURNAL Hereditas 124 (2), 111-120 (1996)
 MEDLINE 96376325
 REFERENCE 2 (bases 1 to 852)
 AUTHORS Escher, S.A.
 TITLE Direct Submission
 JOURNAL Submitted (22-MAR-1996) S.A. Escher, Umea University, Department of
 Genetics, S-901 87 Umea, SWEDEN
 COMMENT Related sequences: M35889, M35688 and X07697.
 FEATURES
 source
 1..852
 /organism="Drosophila simulans"
 /strain="Barcelona"
 /db_xref="taxon:7240"
 /chromosome="X"
 /map="7C"
 repeat_region 1..852
 /rpt_type="TANDEM"
 gene 1..852
 /gene="dec-1"
 exon <1..>852
 /gene="dec-1"
 /number=2
 BASE COUNT 324 a 195 c 232 g 101 t

ORIGIN

Query Match 68.0%; Score 17; DB 6; Length 852;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ttgctgacatcgtgt 18
 |||||||
 Db 237 TTGCTGACATCTGCTGT 221

RESULT 27

DMECTRL/c

LOCUS D.melanogaster DNA from repeated region of dec-1 locus, strain
 DEFINITION Samarkand.
 X96926
 X96926.1 GI:1262758
 VERSION X96926.1 GI:1262758
 KEYWORDS dec-1 locus; tandem repeat.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 930)
 Escher, S.A. and Lambertson, A.
 TITLE Evolution of the dec-1 eggshell locus in Drosophila. III. Sequence
 comparisons of the simulans complex repeated domain reveal
 non-concerted evolution
 Hereditas 124 (2), 111-120 (1996)

JOURNAL

MEDLINE
 96376325
 2 (bases 1 to 930)
 Escher, S.A.
 TITLE Direct Submission
 JOURNAL Submitted (22-MAR-1996) S.A. Escher, Umea University, Department of
 Genetics, S-901 87 Umea, SWEDEN
 COMMENT Related sequences: M35889, M35688 and X07697.
 FEATURES Location/Qualifiers

FEATURES

source

1..930
 /organism="Drosophila melanogaster"
 /strain="Samarkand"
 /db_xref="taxon:7227"
 /chromosome="X"
 /map="7C"
 repeat_region
 1..930
 /rpt_type="TANDEM"
 gene
 1..930
 /gene="dec-1"
 <1..>930
 exon
 /gene="dec-1"
 /number=2

BASE COUNT 352 a 213 c 254 g 111 t
 ORIGIN

Query Match 68.0%; Score 17; DB 6; Length 930;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ttgctgacatcgtgt 18
 |||||||
 Db 393 TTGCTGACATCTGCTGT 377

RESULT 28

DRODECIA/c

LOCUS DRODECIA 3741 bp mRNA INV 28-NOV-2000
 DEFINITION Drosophila melanogaster defective chorion-1 fcl25 protein precursor
 (dec-1) mRNA, complete cds.
 M35887 M35688
 VERSION M35887.2 GI:11386620
 KEYWORDS dec-1 protein; eggshell protein.

Query Match 68.0%; Score 17; DB 6; Length 3741;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

SOURCE

fruit fly.

Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 78)
 Hawley, R.J. and Waring, G.L.
 TITLE Cloning and analysis of the dec-1 female-sterile locus, a gene
 required for proper assembly of the Drosophila eggshell
 JOURNAL Genes Dev. 2 (3), 341-349 (1988)
 MEDLINE
 88243015
 PUBMED
 3378704

REFERENCE

2 (bases 1 to 3741)
 Waring, G.L., Hawley, R.J. and Schoenfeld, T.
 TITLE Multiple proteins are produced from the dec-1 eggshell gene in
 Drosophila by alternative RNA splicing and proteolytic cleavage
 events
 Dev. Biol. 142 (1), 1-12 (1990)

JOURNAL
 MEDLINE
 91032553
 PUBMED
 16998261

COMMENT

On Nov 28, 2000 this sequence version replaced gi:157181.
 Draft entry and computer-readable sequence kindly submitted by
 G.L. Waring, 28-JUN-1990.
 Sequence update by GenBank staff to remove sequence derived from
 genomic DNA, which can be found in X07697.

FEATURES

source

1..3741
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="X"
 /map="7C"
 /tissue-type="ovary"
 1..3741
 /gene="dec-1"
 76..3447
 /gene="dec-1"
 /note="alternatively spliced"
 /codon_start=1
 /product="defective chorion-1 fcl25 protein precursor"
 /protein_id="AA28446.1"
 /db_xref="GI:157182"
 /translation="MRLFSILPLALLVAVAGSRYVSDPATDAGSTTNSPTDKP
 RIPSODEILGOMPSINPIRTGNPOMDAFPMPLAGSLPLKMSLPLPAYSLIGALIDNV
 OPTAASKVYVILADATATKTRNAPNPPLGOMLPLKMSLPLPAYSLIGALIDNV
 SFLAOLPAMPYVPGILGAAPVAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP
 EALQNAFTFPNPNFNDASSLIGSYPTTAPPNLDVYVOMOPFPPTPPAPAPAGT
 DAQASDISEVRVPEDEPYSOEAKMTKSALEMOERKQDAQVQKDEQVPLFRPPT
 ONODATEERTLEHLRVEAKLRAFERQVRLVLEKQRAKRAKNSGTORANALAKRQAS
 ISYPLSRPTIHKITRADIEQALRDYVRLVLEKQRAKRAKNSGTORANALAKRQAS
 ODQTLSEKEDIVQIMAYAYRMANOMSEKQKDYAAAYARTQONPMMOOROMSEEOA
 KIOONOOOIOONPMMOOROMSEEOAKIOONOOOIOONPMMOOROMSEEOAKIOON
 OQIOONPMMOOROMSEEOAKIOONOOOIOONPMMOOROMSEEOAKIOONOOOIOON
 PMMOOROMSEEOAKIOONOOOIOONPMMOOROMSEEOAKIOONOOOIOONPMMOOR
 PMMOOROMSEEOAKIOONOOOIOONPMMOOROMSEEOAKIOONOOOIOONPMMOOR
 ROMSEEOAKIOONOOOIOONPMMOOROMSEEOAKIOONOOOIOONPMMOOROMSEEO
 MAENPOSVOOQPMMOOROMSEEOAKIOONOOOIOONPMMOOROMSEEOAKIOONOOO
 SRCPGSROAOEVOVQGAATVINYVAPVRVVRVTEQATAGGAGSNAGVYRPP
 VNSYSGSGYRAVAGNDEVDMLROHTMAKRTINPQPGVGGSEKQSNSTPTLTP
 APDEQOEHRVHKSPSASPTETENASPSDPQVSGITFYGGILHPPGILLPVRPD
 DPVNNQPYDPNHPHLYTGGGSDYAYLDGHRHRTIMGGGTGHTGILTFQEMERLRLIK
 MFOFRPPPIQGYASPEYSYPCRTAGSGODP"

BASE COUNT

1091 a 987 c 993 g 670 t

ORIGIN

sig_peptide
 mat_peptide

/gene="dec-1"
 /product="defective chorion-1 fcl25 protein"

Query Match 68.0%; Score 17; DB 6; Length 3741;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	2	ttgctgagatctgctgtt 18
Db	1944	TTGCTGATCTGCTGT 1928
RESULT	29	
LOCUS	DRODEC1B/C	
DEFINITION	DRODEC1B 3755 bp mRNA	INV 28-NOV-2000
ACCESSION	Drosophila melanogaster defective chorion-1 fc106 protein precursor (dec-1) mRNA, complete cds.	
VERSION	M35888	M35688
KEYWORDS	M35888.2 GI:11386621	
SOURCE	dec-1 protein; eggshell protein.	
ORGANISM	fruit fly.	
REFERENCE	Drosophila melanogaster	
AUTHORS	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
TITLE	1 (bases 1 to 78)	
JOURNAL	Hawley, R.J. and Waring, G.L.	
MEDLINE	Cloning and analysis of the dec-1 female-sterile locus, a gene required for proper assembly of the Drosophila eggshell	
PUBMED	Genes Dev. 2 (3), 341-349 (1988)	
REFERENCE	88243015	
AUTHORS	3378704	
TITLE	2 (bases 1 to 3755)	
JOURNAL	Waring, G.L., Hawley, R.J. and Schoenfeld, T.	
MEDLINE	Multiple proteins are produced from the dec-1 eggshell gene in Drosophila by alternative RNA splicing and proteolytic cleavage events	
PUBMED	Dev. Biol. 142 (1), 1-12 (1990)	
REFERENCE	91032553	
AUTHORS	1699826	
TITLE	On Nov 28, 2000 this sequence version replaced gi:157183.	
JOURNAL	Draft entry and computer-readable sequence kindly submitted by G.L.Waring, 28-JUN-1990.	
MEDLINE	Sequence update by Genbank staff to remove sequence derived from genomic DNA. which can be found in X07697.	
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		

mat_peptide	/gene="dec-1" 145..2925 /gene="dec-1" /product="defective chorion-1 fcl106 protein"
BASE COUNT	1094 a 991 c 997 g 673 t
ORIGIN	
Query Match	68.0%; Score 17; DB 6; Length 3755;
Best Local Similarity	100.0%; Pred. No. 31;
Matches 17: Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 2	tctcgtcatctctagt 18
Db 1944	TTGCTGGATCTCTCTT 1928
RESULT 30	
DRODEC1C/C	DRODEC1C 5697 bp mRNA INV 28-NOV-2000
LOCUS	Drosophila melanogaster defective chorion-1 fcl177 protein precursor (dec-1) mRNA, complete cds.
DEFINITION	
ACCESSION	M35889 M35688
VERSION	M35889.2 GI:11386622
KEYWORDS	dec-1 protein; eggshell protein. fruit fly.
SOURCE	Drosophila melanogaster
ORGANISM	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 76) Hawley,R.J. and Waring,G.L. Cloning and analysis of the dec-1 female-sterile locus, a gene required for proper assembly of the Drosophila eggshell
AUTHORS	Genes Dev. 2 (3), 341-349 (1988)
TITLE	88243015 3378704
JOURNAL	2 (bases 1 to 5697) Waring,G.L., Hawley,R.J. and Schoenfeld,T. Multiple proteins are produced from the dec-1 eggshell gene in Drosophila by alternative RNA splicing and proteolytic cleavage events
MEDLINE	Dev. Biol. 142 (1), 1-12 (1990)
PUBMED	91032553
COMMENT	1699826 On Nov 28, 2000 this sequence version replaced gi:157185. Draft entry and computer-readable sequence kindly submitted by G.L.Waring, 28-JUN-1990. Sequence update by Genbank staff to remove sequence derived from genomic DNA, which can be found in X07697. Location/Qualifiers 1..5697 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /chromosome="X" /map="7C" /tissue_type="ovary" 1..5697 /gene="dec-1" 76..4845 /gene="dec-1" /note="alternatively spliced" /codon_start=1 /product="defective chorion-1 fcl177 protein precursor" /protein_id="AAA28448.1" /db_xref="GI:157186" /translation="MRLFSLLPLALILVAVGSEVTSDDPATDAGSTTNGTTDRK RIPDEAILGOMPISINIRLTFNGPOMDAFYVMFALGSLKMGSLFPAVSLGAIPLNL OPTKAAAPVTVLADDAATKTRVARARQPPNPIGOLMNPALPDRLPEMDLPGVQ SLPAOLPAMPVTVLIGAAATKTRVARARQPPNPIGOLMNPALPDRLPEMDLPGVQ EALAIQNATFTFNPNANFPAASSLIGQSVTFPPNLDIVYAKORQFFGEMTPAQAAGT DAQASDLSERVRPEDEYSEDAQRKIKSALEMDERQQAQVADQDEVPILMFMPPT QNDQATEKRLLEHLRYEAKLRABERQVIATLRMLQKIELMAKMRSTAAQNDSPVR

ISYPLSTRTPIHKTIRADIEQALRDDYVYRLVKNKEORRANSIGINTOKANLKRQKS
 ODQTLKSDYIOIMAYAVARMANEMESKGGQDXYAYATREOPPMQOROMSEBOA
 KIQONQDQTOONPMQOROMSEBOAKIQONQDQTOONPMQOROMSEBOAKIQONQ
 QOIQONPMQOROMSEBOAKIQONQDQTOONPMQOROMSEBOAKIQONQDQTOON
 PMQOROMSEBOAKIQONQDQTOONPMQOROMSEBOAKIQONQDQTOONPMQOR
 PMQOROMSEBOAKIQONQDQTOONPMQOROMSEBOAKIQONQDQTOONPMQOR
 ROMSEBOAKIQONQDQTOONPMQOROMSEBOAKIQONQDQTOONPMQOROM
 MAENPOGVOOQGMQOMQOMQOMQOMQOMQOMQOMQOMQOMQOMQOMQOMQOM
 SRCSGRQOQOEVQOYGATVINYTAASVYRFRTEATREAGGIGSNAGVPP
 VNSTQSGTAAVGNDEVDMLRQHQTMARTINKRQPEVGGSSQSNSTPTTLP
 APQBPQEHVAKRLAHFRQAGLQATLTKGCGGLDCLGRCRCGRGRLER
 VVSTSTGTCQCKASHRNCREYGLLEIDEGSLNLRKELKLEITISPDYPAE
 ALMRYNAASIREALEPASMCELEIGDOVEDAOEPMEELOHDPNTPOYHNF
 VRLTSTSPITSTTEATPTGSDSTSEATVPTVTTTSTSTSTSTSTSTSTSTST
 ODSOAESSHVTKSTSKQKXIHQHSYVELKLNELKLNELKLNELKLNELKLNEL
 EKNPVVEEPKQEDKREKERVIAEDAPVEPEDEEDSDTSISTTETSPSS
 YSTKPGSLGSPRVDEQSSSNKLDYDDNMWQITLANKRGDYLKSHEROFAQO
 NLEMPKNCNDYNGSOEYGPPEPQADEPSTDEGKAKRALSVQOQLMLNDG
 SDSDGTITTTTTPSPYAMRGKFRRRRSTARVPKPKGASDEWVWSPROAKPDP
 KSSMSKPKOSSOVTQATVSTKLSIDVLKDLVRLQIQKEKSSSLRTGSSNNLSK
 TKPSIKPVKVIKRRRLRRRHSIATTIRSPIDTKA"

sig_peptide
 /gene="dec-1"
 76..144
 /gene="dec-1"
 145..1842
 /gene="dec-1"

mat_peptide
 /product="defective chorion-1 fcl177 protein"

BASE COUNT 1704 a 1469 c 1485 g 1038 t 1 others

ORIGIN

Query Match 68.0%; Score 17; DB 6; Length 5697;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ttgcgtgactgcgtt 18
 ||||||||||||||||
 Db 1944 TTGCTGATCTGCTGT 1928

RESULT 31
 AC017995/c
 LOCUS
 DEFINITION
 AC017995 33204 bp DNA HTG 09-DEC-1999
 Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, In ordered
 pieces.
 AC017995
 AC017995.1 GI:6553195
 HTG; HTGS_PHASE2.
 fruit fly
 ORGANISM
 Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 33204)
 Adams, M. and Venter, J.C.
 Direct Submission
 Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 COMMENT
 This sequence was identified as CDM:10212937 by the submitter.
 For more information on this record e-mail to fly@celera.com.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 FEATURES
 source
 1..33204
 Location/Qualifiers
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 BASE COUNT 8981 a 7720 c 7703 g 8800 t
 ORIGIN

Query Match 68.0%; Score 17; DB 65; Length 33204;
 Best Local Similarity 100.0%; Pred. No. 25;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 ttgcgtgactgcgtt 18
 ||||||||||||||||
 Db 15905 TTGCTGATCTGCTGT 15889

RESULT 32
 AC016952/c
 LOCUS
 DEFINITION
 AC016952 161540 bp DNA HTG 27-AUG-2000
 Homo sapiens chromosome 3 clone RP11-220014, WORKING DRAFT
 SEQUENCE, 15 unordered pieces.
 AC016952
 AC016952.11 GI:9929556
 HTG; HTGS_PHASE1; HTGS_DRAFT.
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 161540)

REFERENCE
 AUTHORS
 Muzny, D.M., Adams, C., Bailey, M., Barbier, J., Blankenburg, K.,
 Bodola, B., Bouck, J., Bowler, S., Brooks, A., Bunay, C., Bunac, C.,
 Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,
 David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N.,
 Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,
 Forcum-Tansey, J., Frantz, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L.,
 Guerra, W., Harris, K., Hernandez, J., Hodgson, A., Hognes, M.,
 Hollaway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, T., Jones, M.,
 Kelly, S., Kondejewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,
 Licharge, O., Liu, J., Liu, W., Logan, O., Lozado, R.J., Lu, J.,
 Lucier, R., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M.,
 Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S.,
 Oswal, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L.,
 Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S.,
 Shieh, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sugan, R.,
 Tabor, P., Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Wabnah, M.,
 Watlington, S., Weinstein, G., Weinstein, I.R., Williamson, A.,
 Worley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Nelson, D. and
 Gibbs, R.
 Direct Submission
 Submitted (09-DEC-1999) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Aug 26, 2000 this sequence version replaced gi:9719634.
 COMMENT
 Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: HMUN
 Center clone name: RP11-220014
 Summary Statistics
 Sequencing vector: M13; L08821
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 138744 bases at least Q40
 Consensus quality: 149435 bases at least Q30
 Consensus quality: 153457 bases at least Q20
 Estimated insert size: 155950; sum-of-contigs estimation
 Estimated insert size: 211269; agarose-fp estimation
 Quality coverage: 3.5x in Q20 bases; sum-of-contigs estimation
 Quality coverage: 4.8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 15 contigs. The true order of the pieces
 * is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 5164: contig of 5164 bp in length
 * 5165 5264: gap of unknown length
 * 5265 55000: contig of 49736 bp in length
 * 55001 55100: gap of unknown length
 * 55101 84235: contig of 29135 bp in length
 * 84236 84335: gap of unknown length
 * 84336 104093: contig of 19758 bp in length
 * 104094 104193: gap of unknown length
 * 104194 117951: contig of 13758 bp in length
 * 117952 118051: gap of unknown length
 * 118052 126732: contig of 8661 bp in length
 * 126733 126832: gap of unknown length
 * 126833 135507: contig of 8675 bp in length
 * 135508 135607: gap of unknown length
 * 135608 141357: contig of 5750 bp in length
 * 141358 141457: gap of unknown length
 * 141458 145691: contig of 4234 bp in length
 * 145692 145791: gap of unknown length
 * 145792 149739: contig of 3948 bp in length
 * 149740 149839: gap of unknown length
 * 149840 152725: contig of 2886 bp in length
 * 152726 152825: gap of unknown length
 * 152826 155298: contig of 2473 bp in length
 * 155299 155398: gap of unknown length
 * 155399 158289: contig of 2891 bp in length
 * 158290 158389: gap of unknown length
 * 158390 160296: contig of 1907 bp in length
 * 160297 160396: gap of unknown length
 * 160397 161340: contig of 1144 bp in length.
 Location/Qualifiers
 1. 161540
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clone="RP11-220014"

BASE COUNT 53135 a 28040 c 27770 g 51176 t 1419 others
 ORIGIN

Query Match 68.0%; Score 17; DB 64; Length 161540;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 3 tgcctgacatcgcgttt 19
 ||||||||||||||||
 Db 22839 TCCTGCAATCTGCTGTTT 22823

RESULT 33
 AEO03442
 LOCUS
 DEFINITION Drosophila melanogaster genomic scaffold 142000013386054 section 26
 of 35, complete sequence.
 AEO03442 AEO02566
 ACCESSION
 VERSION AEO03442.1 GI:7290831
 KEYWORDS
 SOURCE HTG.
 ORGANISM
 fruit fly.
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 306267)
 Adams,M.D., Celinker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
 Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,
 George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
 Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Zheng,X.H.,
 Brandon,R.C., Rogers,Y.H., Blasej,R.G., Champe,M., Pfeiffer,B.D.,
 Man,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabot

TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 FEATURES
 source
 Location/Qualifiers
 1. 306267
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="X"
 join(32686..32823,32937..33134,33225..33387,34597..34762,
 35618..35749)
 /gene="CG1402"
 /note="Partial transcript."
 /product="CT3206"
 /db_xref="FLYBASE:FBan0001402"
 /db_xref="FLYBASE:FBgn0029962"
 <32686..35749
 /gene="CG1402"
 /map="7C3-7C4"
 /db_xref="FLYBASE:FBan0001402"
 /db_xref="FLYBASE:FBgn0029962"
 join(33264..33387,34597..34762,35618..35639)
 /gene="CG1402"
 /note="CG1402 gene product; Partial transcript"
 /codon_start=1
 /db_xref="FLYBASE:FBan0001402"
 /db_xref="FLYBASE:FBgn0029962"
 /protein_id="AAFA6274.1"
 /db_xref="GI:7290832"
 /translation="MSFAQKSGIVGLSIMVOIGENPNPELRITSTFNKVLVGRGS
 TPRIHISVRLSPWTIDHTYIEGSTTHPGCMESVWLIIVKPIITITQEVSLCTNSD
 D"

mrna
 complement(48924..350117)
 /gene="CG10920"
 /product="CT30589"
 /db_xref="FLYBASE:FBan0010920"
 /db_xref="FLYBASE:FBgn0029963"

gene
complement(<48924. .>50117)
/gene="CG10920"
/map="7C3-7C5"
/db_xref="FLYBASE:FBan0010920"
/db_xref="FLYBASE:FBgn0029963"
complement(<48924. .50033)
/gene="CG10920"
/note="CG10920 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0010920"
/db_xref="FLYBASE:FBgn0029963"
/protein_id="AAF46275.1"
/db_xref="GI:7290833"
/translation="MKNPRLPLVAOHARSCAMVELADVPSRAMSTNCOQPIRDO
LRSTEPSGSGSELVWASSLSKTIQENHRPVLQPCRLGYNLLPYEAKVLIO
LHREPLQAPFEMRLIORRPLFLPSYHRVYOHIOHIDGTMGRGLARLAASVDY
LDGLDILATLAPYKRGKRGKSLKEFWMINIIRSLRLRTTYVTYTHPEYVYMRQIA
QYRGSHVTEGIVGSLMPLAQOEGCAGLEPAGWVPRLLDMSVLTFTSLSHICRLPL
MSELQIYNTAVIQMMASLVAYPEVTSICMAGTAPLACPEPMPPLYNMHVLDLTD
LVARGQNGALTEFMRTPKIQLLRSKDNVYRST"
<54229. .>54681
/gene="CG1409"
/product="CT3292"
/db_xref="FLYBASE:FBan0001409"
/db_xref="FLYBASE:FBgn0029964"
<54229. .>54681
/gene="CG1409"
/map="7C4-7C5"
/db_xref="FLYBASE:FBan0001409"
/db_xref="FLYBASE:FBgn0029964"
54229. 54681
/gene="CG1409 gene product"
/note="CG1409 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0001409"
/db_xref="FLYBASE:FBgn0029964"
/protein_id="AAF46276.1"
/db_xref="GI:7290834"
/translation="MARTLAQIILIGAOIVGHALVYKTKROELQAFEDARILQETLKAN
DINSRSAAKTMTLAEAOQILIDVSLDINRQALIDHYOHLFRVNDKSTGGSFYIOSKY
FAKSRIDDELRTLLVKTIDDSHSILITPPESSQFQENKEPQGSK"
complement(<60589. 60716,60988. 62089,62143. .>62916)
/gene="CG15326"
/product="CT35326"
/db_xref="FLYBASE:FBan0015326"
/db_xref="FLYBASE:FBgn0029965"
/evidence=not_experimental
complement(<60589. .>62916)
/gene="CG15326"
/map="7C4-7C5"
/db_xref="FLYBASE:FBan0015326"
/db_xref="FLYBASE:FBgn0029965"
/evidence=not_experimental
complement(<60589. 60716,60988. 62089,62143. 62916)
/gene="CG15326 gene product"
/note="CG15326 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0015326"
/db_xref="FLYBASE:FBgn0029965"
/evidence=not_experimental
/protein_id="AAF46277.1"
/db_xref="GI:7290835"
/translation="MESSSDMDLAEALAQOVVANSFMRGFKTLTYTHTNSOSTGHL
ELDDIYVSNIPNIOARLLILOOSMEKYRYVHAVALIADVIGLPSISATYARARATODS
HTLIYMSPTAYCEBMOATLRLMRISLVWGVYLRPGDHILLVSYFFPSALHGCO
VISANVNRVYQGTKRASQDTPFSKIGNFYGLTCTATWEDMPLVWRPDSGSPG
TEGALLQENAEINFTGLYWNKNEVLATDESGRLFDELFGHADFSLGGRPKRS
AGSELPSYQSTLYFEMSHIMLTINLSASAEKLSFPTPLLMRAIGIVLTLALLLM
LLVRNRHLEPRNDEYELVLTWGNEDMDWVQRFPSRLVLTTLFEATVILRSGYO
SMTYLLRODTRQNPOTISVLAOHNFITQALAEVNEARILASLPELRPOLYVEGSE
LOSPALAOSSSSARVALITPYEYFGEFRKVVMSRRLHYRERYTQOLAEVYRHH
'SHLVNKLKOIHOAHTHGFLIETHTRQVYSANDEKDESVARITASTSYSLDDIDCPSL
SESSEDOVAAPRONVLSMRELAFLWILLIANIGAVVVELETLIPKIKTKILTRM

mRNA

```

KKSRATATTTSSPTTDFSCSKDKGDGDSWPKCSLLYS"
complement(join(63233..63902,64113..65786,
66185..66447,66503..66920,66994..>69146))
/gene="dec-1"
/note="Nucleotide sequence of the C elegans sequence differs from the published sequence for this transcript."
/db_xref="FLYBASE:FBgn0000427"
/db_xref="FLYBASE:FBgn0000427"
complement(<63233..>69146)
/gene="dec-1"
/note="dec-1"
map="7C4-7C5"
/db_xref="FLYBASE:FBan0002175"
/db_xref="FLYBASE:FBgn0000427"
complement(join(63233..63902,64113..65786,65862..65986,
66185..66447,66503..66854))
/gene="dec-1"
/note="dec-1 gene product; Nucleotide sequence of the C elegans sequence differs from the published sequence for this transcript"
/codon_start=1
/db_xref="FLYBASE:FBan0002175"
/db_xref="FLYBASE:FBgn0000427"
/protein_id="Aaf46278.1"
/db_xref="GI:7290836"
/translatation="MMPFALGSILKMGSLFPAYSIGAIIPDNLOPTAAASKVVLVAD
DATKTRVARONPEPNPLGOLMMNPALPOFOPLSDMSGYOVSFLAOLPAMPVTGLD
LGAAPVPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP
FDASLLGGSYPTTPAPPPNLDPVAOMQOFFPCMTPAOPAAAGCTDAOSDSEVRVPE
DPSISQAKMKTKSALKEWEKODQAQVDQOVPLMFNRHTTONDATEKTLEDR
VEALARFERQVIAELMLQIKTELMAQMSASAANGSDSPRISPLSTPIHKITR
ADIEQLARDFYVRLVNKRAQRARSNGINTQAMALAKROAKSQDDTLSEKDYLQATA
YAYMANEOMESEKQKDYAAARTQONPMNQOMROWSEBAKIQONOOIQONPMNQOMROW
MQOWSEBAKIQONNOOIQONPMNQOMROWSEBAKIQONOOIQONPMNQOMROWSEBAK
SEBAKIQONNOIQONPMNQOMROWSEBAKIQONOOIQONPMNQOMROWSEBAK

```

CDS

gene

Query Match 68.0%; Score 17; DB 4; Length 306267;
Best Local Similarity 100.0%; Pred. No. 21;

Matches	Conservative	Mismatches	Gaps	Indels
17	0	0	0	0

Oy 2 ttgcctgacatcgcctgtt 18
|||||
Db 67205 TTGCCTGACATCGCTGT 67221

RESULT 34
DMDCEICTRS5/C DMDCEICTRS 774 bp DNA INV 25-SEP-1996
DEFINITION D.mauritiana DNA from repeated region of dec-1 locus.
VERSION X96930 X96930.1 GI:1262759
KEYWORDS dec-1 locus; tandem repeat.
SOURCE Drosophila mauritiana.
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 774)
AUTHORS Escher, S.A. and Lambertson, A.
TITLE Evolution of the dec-1 eggshell locus in Drosophila. III. Sequence comparisons of the simulans complex repeated domain reveal non-concerted evolution
JOURNAL Hereditas 124 (2), 111-120 (1996)
MEDLINE 96376325
REFERENCE 2 (bases 1 to 774)
AUTHORS Escher, S.A.
TITLE Direct Submission
JOURNAL Submitted (22-MAR-1996) S.A. Escher, Umea University, Department of Genetics, S-901 87 Umea, SWEDEN
COMMENT Related sequences: M35889, M35688 and X07697.
FEATURES
source location/Qualifiers

/organism="Drosophila mauritiana"
/db_xref="taxon:7226"
/chromosome="X"
/map="7C"
1..774
/rpl_type=TANDEM
1..774
/gene="dec-1"
<1..>774
/gene="dec-1"
/number=2

BASE COUNT 290 a 182 c 212 g 90 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 16; DB 6; Length 774;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ttgctgactcgtctgt 17
|||||
Db 159 TTGCTGATCTGCTGT 144

RESULT 35
AF281080/c 1614 bp mRNA VRT 01-OCT-2000
LOCUS AF281080
DEFINITION Xenopus laevis homeobox transcription factor Nkx2-1 (Nkx2-1) mRNA,
complete cds.
ACCESSION AF281080
VERSION AF281080.1 GI:10442649
KEYWORDS
SOURCE African clawed frog.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 1614)
Small, E.M., Vokes, S.A., Garriock, R.J., Li, D. and Krieg, P.A.
Developmental expression of the xenopus Nkx2-1 and Nkx2-4 genes
Mech. Dev. 96 (2), 259-262 (2000)
20417824
2 (bases 1 to 1614)
Small, E.M., Vokes, S.A. and Krieg, P.A.
Direct Submission
Submitted (21-JUN-2000) Cell Biology and Anatomy, University of
Arizona Health Sciences Center, 1501 N. Campbell Ave., Tucson, AZ
85724, USA

FEATURES
source Location/Qualifiers
1..1614
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/note="expressed in ventral forebrain, lungbud primordia
and thyroid anlage"
1..1614
/gene="Nkx2-1"
162..1205
/note="Nkx2-1"
/gene="thyroid transcription factor; TTF-1; T/ebp; chicken
and mouse Nkx2-1 ortholog; contains homeodomain"
/codon_start=1
/product="homeobox transcription factor Nkx2-1"
/protein_id="AKG17405.1"
/db_xref="GI:10442650"
/translation="MSMSPKHTTPFSVDIISPLSESYKKVAMEGAGLAPLTAAYR
OSVOASQOQMGHNGPVSAAVHMTAGVPLSHHTMGCGCNGNLGNLGMSEPLP
YDTRNSATSGANGANDPRESTISRFMGPSGKMGSLGMSGLVGRKSMPLQA
TPRRKRYLFFSOAGVYELERREKOKYLSAPREHLASMIHITPTQVKIMQNHRYKA
KROADKASQOQITQDNSCCOQSPRRAYVYLKDGKPCQAGSNTPTTAAIOSHOO
TATATITTINGLGPQHSHOTNSAGOSPDLVPSNPSLSLONVTSLSLSSSSDYGS
AASCTITLLNGRTW"

gene
CDS
475 a 407 c 399 g 333 t
BASE COUNT
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 16; DB 8; Length 1614;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 tctgctgttgaagcg 25
|||||
Db 717 TCTGCTGTTTGAAGCG 702

RESULT 36
AF143940
LOCUS AF143940
DEFINITION Arabidopsis thaliana SMW2/SNF2-like protein (DDM1) gene, complete
cds; and tRNA-Glu gene, complete sequence.
ACCESSION AF143940
VERSION AF143940.1 GI:4726078
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 5000)
Jeddeloh, J.A., Stokes, T.L. and Richards, E.J.
Maintenance of genomic methylation requires a SMW2/SNF2-like
protein
Nat. Genet. 22 (1), 94-97 (1999)
99251588
2 (bases 1 to 5000)
Jeddeloh, J.A. and Richards, E.J.
Direct Submission
Submitted (17-APR-1999) Biology, Washington University, One
Brookings Drive, St. Louis, MO 63130, USA
Location/Qualifiers
1..5000
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
join(<535..566,772..850,986..1252,1354..1440,1549..1895,
1976..2165,2251..2426,2559..2625,2703..2892,2975..3070,
3148..3242,3317..3436,3540..3659,3745..3843,3934..4038,
4130..>4354)
/gene="DDM1"
/product="SMW2/SNF2-like protein"
/protein_id="DDM1"
/codon_start=1
/product="SMW2/SNF2-like protein"
/protein_id="AAD28303.1"
/db_xref="GI:4726079"
/translation="MWSLRSKVIIPASEVSDGKTEKDAQSDSPSTVLANEENCERKS
YVVEEELILAKNGDSSLISPAQOEEOILKLADEDEKANNAGSAVAPILNTOFK
LDTTLOTOLYSEFLKEMDITINGLESSOKAEPKRGRRKRAASOYNNTKAR
AVAAKISRKEDGETINSIDTEETRYKLNELCPILTGQLKSTQAKYKWLISMQ
NGLGILLADQDGLKTTQTGFLSHLGNGLDGYLYVIAPLSTLSNFFNFIARTPSI
NALIYHGDKNORDLRRKHMPKTVGPFPVITYSEVAMNDARKILKHYPMKYVVIDE
GHRKLNKCKILRELKHLKMDNKLILGTPLQNNLSLWMSLNFILPDIPTSHDEES
WFDESEKNKNATATEEERKRAQVSKHLGLRPLILRRKCVLEISLPKRELIMWA
TMTQHKRPQEHVIVNNTLFAHIGLNAIRGGGKMLNLVYLIRKNNHNDLLOGOD
GSYLYHPVEEIVGCGKRLLEPILRPLVPLNNKHVILFSQWTLDDIMDYFESEKGE
VCRIDGSKLDERRKQIKDSDEKSSCSITLSTRAGGLDINLPADTCILYSDNRP
OMQIADMRCHRIQOTRPVHYRUSTWOSIETRYLKRAYSKLLEHVIIGOGFOHDER
AKSSTPLEEEDILALKEDEFAEDKLIQTIDISADIDLRLDRDLTTTAGETQAAEA
FPVKGPMGEVVLPSGGMLSLNS"

gene
CDS
variation
785

```

/ gene="DDM1"
/ note="site of ddm1-5 (som8) mutation; replace with 83 bp"
/ replace=""
variation      2384..2385
/ gene="DDM1"
/ note="site of ddm1-6 (som4) mutation"
/ replace=""
misc_feature   3186
/ gene="DDM1"
/ note="alternate splice donor site used in ddm1-2"
variation      3243
/ gene="DDM1"
/ note="site of ddm1-2 mutation"
/ replace=""
variation      3337
/ gene="DDM1"
/ note="site of ddm1-7 (som5) mutation"
/ replace=""
trna           complement(4755..4826)
/ product="tRNA-Glu"
/ evidence="not_experimental"
BASE COUNT    1386 a 909 c 1029 g 1676 t
ORIGIN

Query Match      64.0%; Score 16; DB 13; Length 5000;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3  tctgagctgctgtt 18
Db      11332 tctgagctgctgtt 1147

RESULT 37
LOCUS      YL1012084      6887 bp      DNA      PLN      23-SEP-1999
DEFINITION Yarrowia lipolytica pex10 gene for peroxin 10.
ACCESSION  AJ012084
VERSION     AJ012084.1 GI:5921565
KEYWORDS   peroxin 10; pex10 gene.
SOURCE      Yarrowia lipolytica.
ORGANISM    Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Dipodascaceae; Yarrowia.
            1 (bases 1 to 6887)
REFERENCE   1 Dail,M.T., Galliardin,C. and Nicaud,J.M.J.
            PEX10 is essential for peroxisome biogenesis in Yarrowia lipolytica
            Unpublished
            2 (bases 1 to 6887)
AUTHORS     Nicaud,J.M.
TITLE       Direct Submission
JOURNAL     Submitted (20-SEP-1999) Nicaud J.M., Laboratoire de Genetique des
            Microorganismes, INRA-CNRS, BP 01, Tiversal-Grignon, F78850, FRANCE
FEATURES
            source
            1..6887
            /organism="Yarrowia lipolytica"
            /strain="W29"
            /db_xref="taxon:4952"
            complement(2293..3357)
            /gene="pex10"
            complement(2293..3357)
            /gene="pex10"
            /function="essential for peroxisome biogenesis"
            /codon_start=1
            /evidence="experimental"
            /product="peroxin 10"
            /protein_id="CAB56484.1"
            /db_xref="GI:5921566"
            /translation="MSDNTITKKIRKPIRTERLPYAGAEITIRANOKHYEVSIE
            /OHLYFLQKMGVRFTHQYKELETFASKFALGLCTLVGSKTIGEEYTNLMWYTRDRT
            ALPGVVRFGVYLSNTLPYLFVRYMGKLRKRLMRREYPHLVEYDEDEPSPSPETMKER
            VIKTFVNKDFKTALEGTALHIAIFVYGSYOLSKRIWMRYVFGHRIDKNEPRIG

```

```

YEMLGILIFAFATSPVOTGREYIALLEKSEKGEKDEKAEAVPKKSSIPFIE
DTFGTEDEKIDLEDDPROLKFIEPESACTLCLSTYSAPACTCGHFFCMDCISEVRE
KPCPCRCRCVRSQNLPIR"
BASE COUNT      1793 a 1841 c 1485 g 1768 t
ORIGIN

Query Match      64.0%; Score 16; DB 15; Length 6887;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8  gatctgctgttgaag 23
Db      4051 GATCTGCTGTTGAAG 4066

RESULT 38
AE004322/c      11328 bp      DNA      BCT      31-JUL-2000
LOCUS          Vibrio cholerae chromosome I, section 230 of 251 of the complete
DEFINITION     chromosome.
ACCESSION      AE004322 AE003852
VERSION        AE004322.1 GI:9657104
KEYWORDS       Vibrio cholerae.
SOURCE         Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
ORGANISM       Vibrio cholerae.
REFERENCE       1 (bases 1 to 11328)
                Heideberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwin,M.L.,
                Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.,
                Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
                Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I.,
                Sellers,P., McDonald,L., Uitterback,T., Fleischmann,R.D.,
                Nierman,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R.,
                Mekalanos,J.J., Venter,J.C. and Fraser,C.M.
                DNA sequence of both chromosomes of the cholera pathogen Vibrio
                cholerae
                Nature 406 (6795), 477-483 (2000)
                2 (bases 1 to 11328)
                Heideberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwin,M.L.,
                Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.,
                Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
                Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I.,
                Sellers,P., McDonald,L., Uitterback,T., Fleischmann,R.D.,
                Nierman,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R.,
                Mekalanos,J.J., Venter,J.C. and Fraser,C.M.
                Direct Submission
                Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
                Medical Center Dr., Rockville, MD 20850, USA
FEATURES
            source
            1..11328
            /organism="Vibrio cholerae"
            /strain="N16961"
            /serotype="O1"
            /db_xref="taxon:666"
            /chromosome="I"
            /note="biotype: El Tor"
            complement(95..286)
            /gene="VC2521"
            complement(95..286)
            /gene="VC2521"
            /note="identified by Glimmer2; putative"
            /codon_start=1
            /transl_table=11
            /product="hypothetical protein"
            /protein_id="AAF95663.1"
            /db_xref="GI:9657105"
            /translation="MANEDVYGIKREVEVPINYPPLAILPVCHSKRQPLKVRLS
            SACYRCMMKRTIGFLDK"
            524..1522
            /gene="VC2522"
            524..1522
            /gene="VC2522"

```

/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF95664.1"
/db_xref="GI:9657106"
/translation="MAFIFYYVGINMLI EAVVFLFGLALIVMGADKLIVGSAALAR
NIGISPLVIGMTILAMGSSAPEMVYSAATALEGKTDTAVGNVLSNIALIIGITA
IKPLPSGAVGAVLRELPMVIGVTLIAGAILMDNHLIGFEGILLALFAAPILVMQVS
REKONGALLDEQSEIPIGVSNPKAEMVYVGLILPIGGMIVDNAVYIAKFGM
SGLVIGRTIANGTSLPELASIACALGEDDMANGNIIIGSNVFNILAVMGPGLLNP
SLSPAMSRDWMVGSLILVGLAKKRQITRLLEGALLCSFIIYITLLMNI"
1572..2552
/gene="VC2523"
1572..2552
/gene="VC2523"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF95665.1"
/db_xref="GI:9657107"
/translation="MRPHSEDTQVANAQVATELTHAIQQL EQYIINDPAPACAMILA
NOTGVVVMGKSGHIGKIKIATLASGTSAFVHPPEASHGDLGMIERDYLAIIS
NGSESEILALPLVLRSLIRVISMGNPNSMAKLADIHLQITVPRACPELAPTS
STTATLVMDALVALMOARCTAEDFALSHPGALGRKLLKINDIMHSGDLKVA
POLLRLDLLEISQGLGIMTAIVDEDFLLGIFTDGDLRLIDKRIDITHSVIADVMT
RQPVAPNLLAEBGLNMQAKRINGMLVENVKLVGALNMHDLIKAGVM"
2552..3109
/gene="VC2524"
2552..3109
/gene="VC2524"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF95666.1"
/db_xref="GI:9657108"
/translation="MSTVSTLYGEVPSLLEIAKOIKLLICDVGVSESDGIYMGNO
GELTQFHRDGYKALNMAGIEIATITGRSQVYENRMKALGSLIIQGDQDKVQA
YDITQKLAIAPEOTGYIGDDIIMPVMEKVALRVGADGHPLLAQRANVYTHIKGGH
GAVREVCDDILQARNELDVHKGSLI"
3106..3669
/gene="VC2525"
3106..3669
/gene="VC2525"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF95667.1"
/db_xref="GI:9657109"
/translation="MSLSRIYVLLFLTASLSLYLLLGQEDSKIQVAPNELPMFSG
ENLENI SYDEGIRNVYITSLIHLYAASGNTLFPAPILKYRECTILOEWETARGI
LSKQDVLFLYDVLAKNLLPDSGPDTLTSEMSIOLKSRDEWADKPVELRGPOPEETHG
QAMKGNFADHSELTYNQVGREYELTP"
3605..4144
/gene="VC2527"
3605..4144
/gene="VC2527"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF95668.1"
/db_xref="GI:9657110"
/translation="MPIIRLNFITKRYVDMKRSLSLTAQLASTAQAVALSTREOPY
YIDSQQLDMKSNRPFILGDVCLKQSGININADKLIVIRNAENGKIEIETGYNAVAT
FSQLTDEGKTYLGEAKELIYKRVDELVIKAMLSQDSEIRSKIKYKISLQKLLA
DSDGKRVSTVQLQPAIQE"
4147..4872
/gene="VC2528"
4147..4872

/gene="VC2528"
/note="similar to PID:2317737; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ABC transporter, ATP-binding protein"
/protein_id="AAF95669.1"
/db_xref="GI:9657111"
/translation="MALIKAKHAKSKKKRKYSDVSLQVESQIYVGLLPNAGKTT
SFTIVGLVARDSTITIDNDISILPMHSRSMGIDYLPQEASTFKLSVEONIMAV
LQTEELTHEERQKLEDLLEFPIQHIRKSAGMALSGERRVEIARALANPOFL
LDEPFAGVDPIPSVDKQILIEHLDRGLGLVITDHNRETLDVCEKAYIVSQRLIAE
GTPQDVNLNEQVKVYLGEOGRL"
4928..6391
/gene="VC2529"
4928..6391
/gene="VC2529"
/note="similar to PID:2317738; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma-54 factor"
/protein_id="AAF95670.1"
/db_xref="GI:9657112"
/translation="MKPSLOIKIGQOLAMTPOLOQATRLQLSTLDLQORIOBAPSN
PLIEVENODEAPSLDSVVRMEPSRPREPELYPEPQSDLI EKSISELEMDPTM
DEVYSANTGSGALDDDAPIYQGETTQTIQDYLHMODLTPPSVDRTIAYALLDAI
DDGYLVLSLEIEQESLRSDDIELEAVAKRRIQODPPGVASLNDQDLLQLTTY
PCDTPMLEEARLLISQYIDDLGNRDYPTIILKETRLKEDRELITOLIQODPPGSGRI
AQDHAEVYIPDVSVYKBOGRMLVTINPDSVPLKINQYADMLRGNNAESNYIRTNQ
EAKWLKISLESRNETLLKVAKCIYEHQDFEEGEEAKPMVINDVAMAEHMESTIS
RVTKQKMHPTPGRIFFELKPFSSHVSTDNQGECSSTAIRLIRIKKIVAENPARPLSDS
KIATLADQGIQVARRITAKYRESLGIAPSSQRRLL"
6416..6703
/gene="VC2530"
6416..6703
/gene="VC2530"
/note="similar to PID:2317739; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="sigma-54 modulation protein, putative"
/protein_id="AAF95671.1"
/db_xref="GI:9657113"
/translation="MQINIGHHIDLDSMODVYHSFKYLERPPDHINHYQIYLA
KLQIATLHVNOAETHAHADENMTAIDSLVDKLYROLNKRKLSH"
6706..7152
/gene="VC2531"
6706..7152
/gene="VC2531"
/note="similar to PID:2317740; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="PTS system, nitrogen regulatory IIA component"
/protein_id="AAF95672.1"
/db_xref="GI:9657114"
/translation="MQLSEVLSLDCKRSAAVOCSSKRALETISEIALHTGQNTLEF
ECMLSRKMGSTGIGNGIALIPHARMSDSENAVALVLLQCEPFIEDSIDNRPDLLPAL
LVPEQCKEHLKTLACMAERLNDKLLKOLRMAASODELYINIMVNO"
7165..8025
/gene="VC2532"
7165..8025
/gene="VC2532"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF95673.1"
/db_xref="GI:9657115"
/translation="MRLIVVSGSGAGKSVALLVLELDGYCYVDNLPSILTAFLQSV
QGSQONVAVSIDIRNLPEKPSLVQVDVLDQIKONNDVSMFLDASKETLLKRYSETRRI

Query Match 64.0%; Score 16; DB 1; Length 11328;
 Best Local Similarity 100.0%; Pred. No. 97;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 tggatctgctgttga 21
 ||||||||||||||||
 Db 7838 TGGATCTGCTGTTTGA 7823

RESULT 39
 AE004026 13856 bp DNA BCT 06-JUL-2000
 LOCUS Xylella fastidiosa, section 172 of 229 of the complete genome.
 DEFINITION AE004026 AE003849
 ACCESSION AE004026.1 GI:9107228
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM Xylella fastidiosa.
 Xylella fastidiosa
 Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 Xylella.

REFERENCE 1 (bases 1 to 13856)
 AUTHORS Silvestri, M.L., Siqueira, W.J., de Souza, A.A., de Souza, A.P.,
 Terezi, M.F., Truffi, D., Tsai, S.M., Tsubako, M.H., Vallada, H., Van
 Slyks, M.A., Verjovski-Almeida, S., Vettore, A.L., Zago, M.A., Zatz, M.,
 Meidanis, J., and Setubal, J.C.
 The genome sequence of the plant pathogen Xylella fastidiosa. The
 Xylella fastidiosa Consortium of the Organization for Nucleotide
 Sequencing and Analysis, Sao Paulo, Brazil
 Nature 406 (6792), 151-157 (2000)
 20365717
 2 (bases 1 to 13856)
 AUTHORS Simpson, A.J.G., Reinach, F.C., Arruda, P., Abreu, F.A., Acencio, M.,
 Alvarenga, R., Alves, L.M.C., Araya, J.E., Bala, G.S., Baptista, C.S.,
 Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bove, J.M., Briones, M.R.S.,
 Bueno, M.R.P., Camargo, A.A., Camargo, L.E.A., Carraro, D.M.,
 Carrier, H., Colauto, N.B., Colombo, C., Costa, F.F., Costa, M.C.R.,
 Costa-Neto, C.M., Coutinho, L.L., Cristofani, M., Dias-Neto, E.,
 Docena, C., El-Dorri, H., Facincani, A.P., Ferreira, A.J.S.,
 Ferreira, V.C.A., Ferro, J.A., Fraga, J.S., Franca, S.C., Franco, M.C.,
 Frohme, M., Furlan, L.R., Garnier, M., Goldman, G.H., Goldman, M.H.S.,
 Gomes, S.L., Gruber, A., Ho, P.L., Hohensei, J.D., Junqueira, M.L.,
 Kemper, E.L., Kitajima, J.P., Krieger, J.E., Kuramae, E.E., Laigret, F.,
 Lambais, M.R., Leite, L.C.C., Lemos, E.G.M., Lemos, M.V.F., Lopes, S.A.,
 Lopes, C.R., Machado, J.A., Machado, M.A., Madeira, A.M.B.N.,
 Madeira, H.M.F., Marino, C.L., Marques, M.V., Martins, E.A.L.,
 Martins, E.M.F., Matsukuma, A.Y., Menck, C.F.M., Miracca, E.C.,
 Miyaki, C.T., Monteiro-Vitorello, C.B., Moon, D.H., Nagai, M.A.,
 Nascimento, A.L.T.O., Netto, L.E.S., Nhani Jr., A., Nobrega, F.G.,
 Nunes, L.R., Oliveira, M.A., de Oliveira, M.C., de Oliveira, R.C.,
 Palmieri, D.A., Paris, A., Peixoto, B.R., Pereira, G.A.G., Pereira
 Jr., H.A., Pesquero, J.B., Quaggio, R.B., Roberto, P.G., Rodrigues, V.,
 de M. Rosa, A.J., de Rosa Jr., Y.E., de Sa, R.G., Santelli, R.V., Silva
 Sasaki, H.E., da Silva, A.C.R., da Silva, F.R., da Silva, A.M., Silva
 Jr., W.A., da Silveira, J.F., Silvestri, M.L.Z., Siqueira, W.J., de
 Souza, A.A., de Souza, A.P., Terezi, M.F., Truffi, D., Tsai, S.M.,
 Tsubako, M.H., Vallada, H., Van Slyks, M.A., Verjovski-Almeida, S.,
 Vettore, A.L., Zago, M.A., Zatz, M., Meidanis, J., and Setubal, J.C.
 Direct Submission

TITLE Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
 Journal Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
 13083-970, Brazil

REMARK
 FEATURES
 source

gene
 CDS

Location/Qualifiers
 1..13856
 /organism="Xylella fastidiosa"
 /db_xref="taxon:2371"
 complement(195..803)
 /gene="XF2101"
 complement(195..803)
 /gene="XF2101"
 /note="similar to GI1652989 (percent identity: 30 %/query
 alignment coverage: 87.6 %/subject alignment coverage:
 67.6 %); identified by sequence similarity; putative; ORF

gene
 CDS

gene
 CDS

gene

CDS

gene
 CDS

gene

CDS

located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA084900.1"
 /db_xref="GI:9107229"
 /translation="MSRCITGVYRWQWQICVYVLPFVAVAGMIVAGIGORDQASS
 DVIIVGAAAYDAKPSPEFERRRHLVYKQGYAPTLITGSGRGSSARPAESQVYR
 RIALRYDPPRALITVSRTTRQNLQARQLMRVYGLRATVYSDPLHMAERLQCR
 ELIDIALASAPSTPRSLHTRWRFLLQVYEFHDLVLIQAN"
 /gene="XF2102"
 complement(800..1042)
 /gene="XF2102"
 complement(800..1042)
 /note="hypothetical protein; identified by sequence
 similarity; putative; ORF located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA084901.1"
 /db_xref="GI:9107230"
 /translation="MARVIMFVSSPALSVKSCITGVYTCMLLHAYVFCGCGERRAL
 CCVAMVLCVHDISARLSGCGSAGFCVSVLVVS"
 /gene="XF2103"
 1397..1732
 /note="hypothetical protein; identified by sequence
 similarity; putative; ORF located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA084902.1"
 /db_xref="GI:9107231"
 /translation="MKARLFLPLMSAISCATYVTSNGYHGPVVEYHPPVYV
 PHYGVHDFENPPIPRDPRYRTHIHLPYQRPDHPDSRPLPKOPKQOPNHE
 MGRDKOP"
 1914..2024
 /gene="XF2104"
 1914..2024
 /note="hypothetical protein; identified by sequence
 similarity; putative; ORF located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA084903.1"
 /db_xref="GI:9107232"
 /translation="MNADQDMLLCSLELMLKICPHCVTLKHQRLAAR"
 complement(2067..2225)
 /gene="XF2105"
 complement(2067..2225)
 /note="hypothetical protein; identified by sequence
 similarity; putative; ORF located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA084904.1"
 /db_xref="GI:9107233"
 /translation="MAAMSREBAMMEFYVCVIFRNALFRVGFYAVDVVAPAAALRL
 QAQRSGLE"
 2271..4160
 /gene="XF2106"
 2271..4160
 /note="similar to SP1P25756 (percent identity: 65 %/query
 alignment coverage: 98.6 %/subject alignment coverage:
 98.4 %); identified by sequence similarity; putative; ORF
 located using Glimmer/RBSfinder"
 /codon_start=1
 /transl_table=11
 /product="glucose inhibited division protein A"

CDS
 /gene="PA1514"
 complement(2452..2961)
 /gene="PA1514"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA04903.1"
 /db_xref="GI:9947471"
 /translation="MRTKLEPLKEAFAPRGDYIETAGSDYFMINNGSTRYKLA
 VETAPEDMAITSTFSAPKLEMPRLRLEHPLUGDAFIPLEGPFLLVAVPLGDP
 VGLVRAFLINGROGVNHYRGVHHPPVLTTEKRDPLVDRSGSGNCDHEFTEDCO
 LLDPOSNO"
 complement(3016..4014)
 /gene="a1c"
 /note="PA1515"
 complement(3016..4014)
 /gene="a1c"
 /codon_start=1
 /transl_table=11
 /product="allantoinase"
 /protein_id="AA04904.1"
 /db_xref="GI:9947472"
 /translation="MNAHSPFRHYLDLADRLGSGVAVASDEMFAPASRMLOAGPEV
 WKEGFPDSCGMKMGWETERRKRFEGHDQAVIRLGPVGLKGVDDITREFFGNHPAS
 LDGFCVAGEDDDSTWSSEVLAAGLQDSHHYPIIDDERWTHLRNLINYPDGIARL
 RLYGVYKDMGNOPPTALDLAAVNGSRALACSDHGRGMNLINPRAIINMGDWE
 TGRRTTGHDMVIALGHPSIEAAVVDTHFKNGNPESCSIOAFVPEGCNEARIEMO
 SLFRRLIPAKLMEHREHRLNALGPTTHVRLNIFPDGVSRLRLGRPOLP"
 complement(4032..4547)
 /gene="PA1516"
 complement(4032..4547)
 /gene="PA1516"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA04905.1"
 /db_xref="GI:9947473"
 /translation="MSRFOULTPASLSREAFTEAFADIEHSPVVAEKAYDLGIDSL
 NDIEGHORMADILLASREOALINAHPLACKAIRGELTASSTSPDAGGTHSC
 SAEAFRTLENDAYKARFGRPIKAVKSGNRHQLIAFERIORHSADEFATALAI
 NKIALFRLOOL"
 complement(4544..5470)
 /gene="PA1517"
 complement(4544..5470)
 /gene="PA1517"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA04906.1"
 /db_xref="GI:9947474"
 /translation="MSADYPRDLIGYNNPMPHMPGDARIALSFVLYEEGRCVL
 HGKRESEFLSEMYAOLPGVGRMSYIEYSGRAGVBLLEFKRNVPLTFVAV
 AMAORNFENVIRAWVADGHEICSHGYSRMYIYOYVDEAOERHMLAIRITLITLTFV
 VGYTGRGPRTRRLVMEEGELIDSDTYDDDLFYMPASTAEKPHIVITYITDITDM
 RFTVOGFGNNGEORFYLIKDFEVLVEGATPAMLSIGLHRLIGRPAWALIERFI
 OYASDHKWFAREDIRAHWHREHPQETEA"
 5843..6223
 /gene="PA1518"
 5843..6223
 /gene="PA1518"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA04907.1"
 /db_xref="GI:9947475"
 /translation="MSASGRPMPGRLLTHVLDSAHGCPGHGKIKELRYVGGQLELI
 ARTLTNHDRCDDPVLOGPDRGVYOLVFNAGVYIARGVOLPEPAFLDQVYLRGFI
 ASADHNVHPLLISYSTYRGS"
 complement(6464..7813)
 /gene="PA1519"
 complement(6464..7813)
 /gene="PA1519"
 /codon_start=1

/transl_table=11
 /product="probable transporter"
 /protein_id="AA04908.1"
 /db_xref="GI:9947476"
 /translation="MESTKOEQGIYAATPPATGLERLEFKLRQHGTVVPELAGLT
 TPTMAVITFVNENINADAGIDHGAFAVTCALAAAGCTMLGLVNNPVGLAPGGLN
 APTTYVGVGMSVMOIALGAVFISGVMMLLTFSVRBMLNLSIPIRIRFGGILISI
 LFTGLIGKTACIGVAPATLTKIGLTSFGLLAACPTMTANSIEXRFRFGGILISI
 LSTVLVGMALGLVQYGVFSAPPSTAPFLANDIGARVNTVISTILFELVHMDTA
 GTLMGVAQRANHLVKEDEGRILENSKAKKADASSVFGMLGPPVTVYESSAGVAGG
 RGLTAVGVVLEVAAMPFAPLAGMIPATFAGALIVYAMLMGMGMHNDMDHEHETI
 PAIVTVIMPLLEFVSVDGIALGITVYAKVLTGKHREVSASLYALVFAKFTFL"
 8188..8967
 /gene="PA1520"
 8188..8967
 /gene="PA1520"
 /codon_start=1
 /transl_table=11
 /product="probable transcriptional regulator"
 /protein_id="AA04909.1"
 /db_xref="GI:9947477"
 /translation="MTDQLQHTRKPRPGKAARSGTODEIYVAHIFDALLEGRLAPGT
 KISEALGEITFGVSRITIRRALSRLAHEQVLLRNRGAVVAPSIDARQILFARRT
 VERAITELATDNATAERLAEIRMVQESSFARGDRGATLSEFHLKLAEMARNA
 PLVSFORSIVSOSLIIAOYEGSGRHCFSFDEHETLDAIEKDKERAVTLMHMHMEH
 IOSKLMLESNDMSGDLHVAFVSHLIGKKRRAKKADSSAA"
 complement(9021..10325)
 /gene="PA1521"
 complement(9021..10325)
 /gene="PA1521"
 /codon_start=1
 /transl_table=11
 /product="probable guanine deaminase"
 /protein_id="AA04910.1"
 /db_xref="GI:9947478"
 /translation="MTRNAKAYRAALILHSIADPAVGVERSYEFEDGLLVEDKVA
 RLGDAPETLIGEGVEVEYRDALITPGFTIDHIFHPQGMJASYGEOLLDVNLNTPF
 PTERGODAHADOVAEIRFLOELRNGTTFALVFGSVHROSVSEFEARLRDLRLIA
 GVWMBRNAPDVLITDPAESSYRDSKALIERMHGGRLLIYAVTPRAPSTAREOLMAA
 RLIREHPGYLTHHSENKELIEWKELPREPSGLDYVDHIGLLGPPSVARHGHLC
 DEBCORLAEFGSAVAFCEPTSNLFSGSLGFLKLRVYKVGIGTGVACGSFSQDS
 LNEAYKVMQLGARRDPEKSLYLATLGARALELDRIGSFATSMEDAFVLDYHATP
 LLSYRLSQAGSLAERLFLTLITIGDDRTYKTFPAAGRSVHRRD"
 complement(10395..11234)
 /gene="PA1522"
 complement(10395..11234)
 /gene="PA1522"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA04911.1"
 /db_xref="GI:9947479"
 /translation="NMWISALAEIORSRGPVLYVTIIEBGRSTPNNAKSVVSAERL
 YDTIGGHLIETRAQAIARMLAARTODTRLEFSIGASLIGCCGCGATVLLLEPKQPO
 AHIAVAGAHVRAVPLPLASLPCRVNWDSEHEFPALDPGEKVNDVLEVER
 MPBSHFIVTINNHPIDELTAAILARNDFAVFGILGSKTRKVRKEHRLRRGVDAER
 LQMRCPMGLSEQVKIPAEIAISIAGEVIATYNAAFGLERKQGPSVTRILTPSRRA
 HES"
 Query Match 64.0%; Score 16; DB 1; Length 15247;
 Best Local Similarity 100.0%; Pred. No. 95;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 gtcgcgacatctgc 16
 |||||||||||||
 Db 1254 GTTGGTGGATCTGCTG 1269
 RESULT 41
 HS314G4 31557 bp DNA PRI 23-NOV-1999
 LOCUS HS314G4
 DEFINITION Human DNA sequence from cosmid 314G4 from a contig from the tip of
 the short arm of chromosome 16, spanning 2kb of 16p13.3. Contains

ESTs and Cpg islands.
 269667
 VERSION 269667.1 GI:1200312
 KEYWORDS 16p13.3; Cpg island.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidi; Homo.
 1 (bases 1 to 31557)
 REFERENCE
 AUTHORS Thomas K.
 TITLE Direct Submission
 JOURNAL Submitted (22-FEB-1996) Sanger Centre, Hinxton, Cambridgeshire, CB10 1RO, England. E-mail enquires regarding this sequence: humquery@sanger.ac.uk
 COMMENT
 IMPORTANT:
 This sequence is not the entire insert of clone 314G4. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
 The true right end of clone 314G4 is at 31557.
 314G4 is from a 280kb clone contig extending from the telomere of 16p. Higgs D.R., Flint J. unpublished. MRC Molecular Haematology Unit, Institute of Molecular Medicine, Oxford.
 314G4 came from the Los Alamos, Flow sorted human Chromosome 16 libraries constructed by Norman Doggett (unpublished).
 Location/Qualifiers
 1. 31557
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /map="16p13.3"
 /clone="LA16-314G4"
 /clone_1lb="LA16"
 428. 716
 /note="Alu repeat: matches 308. .1 of consensus"
 914. .1691
 /note="expressed region - matched by multiple ESTs"
 1692. .2316
 /partial
 /note="match: 5' EST R51457 clone 39198"
 2290. 2810
 /partial
 /note="match: 5' EST T87094 clone 115315"
 complement(3171. .3720)
 /note="expressed region - matched by multiple ESTs"
 5136. 5489
 /note="3 copies of 118 mer 88 & conserved"
 5908. 6086
 /note="match: EST Z39463 clone c-lag06"
 7536. 7567
 /partial
 /note="Alu repeat: matches 1. .32 of consensus"
 7568. 7856
 /note="Alu repeat: matches 1. .308 of consensus"
 9446. 9734
 /note="Alu repeat: matches 308. .1 of consensus"
 12398. 13777
 /note="putative Cpg island"
 13241. 13275
 /note="7 copies of 5 mer 86 & conserved"
 14561. 14677
 /partial
 /note="Alu repeat: matches 254. .136 of consensus"
 14680. 14971
 /note="Alu repeat: matches 308. .1 of consensus"
 14974. 15106
 /partial
 /note="Alu repeat: matches 145. .1 of consensus"
 15127. 15231
 /partial
 /note="Alu repeat: matches 116. .1 of consensus"
 15248. 15355
 /note="Alu repeat: matches 308. .1 of consensus"

repeat_region 15607. .15917
 /note="Alu repeat: matches 308. .1 of consensus"
 repeat_region 16005. .16114
 /partial
 /note="Alu repeat: matches 108. .1 of consensus"
 repeat_region 16123. .16283
 /partial
 /note="Alu repeat: matches 308. .150 of consensus"
 17661. 18594
 /note="putative Cpg island"
 18912. .19019
 /note="3 copies of 36 mer 89 & conserved"
 19514. .19621
 /note="MER22 element fragment"
 19922. 20499
 /note="match: 5' EST H21176 clone 173392"
 20504. .20539
 /note="4 copies of 9 mer 97 & conserved"
 21472. .21766
 /note="Alu repeat: matches 308. .1 of consensus"
 25040. 25378
 /note="expressed region - matched by multiple ESTs"
 complement(25620. .25910)
 /note="expressed region - matched by multiple ESTs"
 26707. 26817
 /partial
 /note="Alu repeat: matches 298. .184 of consensus"
 26818. 26963
 /partial
 /note="Alu repeat: matches 160. .1 of consensus"
 complement(27039. .27209)
 /note="expressed region - matched by multiple ESTs"
 30474. 30611
 /note="3 copies of 46 mer 94 & conserved"
 31087. .31220
 /note="expressed region - matched by multiple ESTs"
 31496. 31557
 /partial
 /note="Alu repeat: matches 1. .62 of consensus"
 5713 a 9969 c 9918 g 5957 t
 BASE COUNT
 ORIGIN
 Query Match 64.0%; Score 16; DB 92; Length 31557;
 Best Local Similarity 100.0%; Pred. No. 89;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 gtgtcgtgattctgtc 16
 |||
 Db 10104 GTTGCTGCATCTGCG 10119
 RESULT 42
 SC8520X 41200 bp DNA 11-AUG-1997
 LOCUS S.cerevisiae chromosome XIII cosmid 8520.
 DEFINITION Z49705 Z49700 Z71257
 ACCESSION Z49705.1 GI:825356
 VERSION
 KEYWORDS ALD2; aldehyde dehydrogenase; ATPase; CBF3B; centromere DNA-binding protein; coiled-coil; DNA mismatch repair; DnaJ; Mitochondrial carrier protein; MHL1; SMP2; transfer RNA-Met; transfer RNA-Val.
 SOURCE
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 REFERENCE
 1 (bases 1 to 41200)
 HUNT, S. and Bowman, S.
 JOURNAL Unpublished
 2 (bases 1 to 41200)
 Barrell, B. and Rajandream, M.A.
 AUTHORS Direct Submission
 TITLE Submitted (19-MAY-1995) Saccharomyces cerevisiae chromosome XIII
 JOURNAL sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge

CDS
11387..13837
/note="YM8520.09, unknown, len: 816, CAI: 0.13"
/codon_start=1
/product="unknown"
/protein_id="CA89796.1"
/db_xref="GI:825565"
/db_xref="SWISS-PROT:Q03823"
/translation="MDVVOOKODKOLOHOTOFOOIREDOCEVPPORPROONRMKPMW
NSTADDEPTGRIAEYINGNGRSGSPPTDPODSDNSNNKGIWMSIASFATSRYSAP
LYVDNTRYSQNTBOINFLNEAKDIISSKSKSCWYELPHISNSSNIDSIDTPG
IISVSGTSGACPLPLNKTPGSGNGVNFINDSLILPSDPLPLHVPDLRTKVLN
TIKNYNEPNEQHLROKTKALKDKRIIISVVDLPKEKESLESORAYVLSR
KLSQNLAEOPORVLTLSFOCLHMODLIPYKCEVEILNHMAHLFEKVDSTFFEVGY
HSVPLTLLAKYIVONNEYLEEDENTVYSVLSFOGLOOYRPMDSSTPTSSYNLG
SNSSTENDSNDSNDNPTTSQOIKKEQLFOGIDKROODTSLKIKVRRIDSSEK
LYQDALDLFLFWMDTFRLLTFRCKLDNERTISEKLAIDYNEKILRNLMCKRYGID
LKNANNLNDTDEATSNINDVHVRPNVESLKIPTNNLEFITLMDLMTTENIGYK
QPIPLINLSPEFISRSFNDYTLPPNIRKOYONSNNKIMLOEDMSKMNKHQNLNDQR
EGESLSSSESLIPENISTVKDFLOFVOYONKSSDFVATIDYDDKVKCYKFLYNT
IPTKNPLSRKHLRLNIDTPTSLINTVQYOLWVKIHDSFKLIQLKNLDPREIPHA
LRLSTLNCFLDSTSTSGPVRQROTVELRLRLTETWRTYQDWSPTGKIKHLRDLIS
VLAYNDNPKNLINDVRR"
14039..14713
/note="YM8520.10, putative DNAJ gene, len: 224, CAI: 0.14,
similar to DNAJ proteins"
/codon_start=1
/product="unknown"
/protein_id="CA89797.1"
/db_xref="GI:825566"
/db_xref="SWISS-PROT:P48353"
/translation="MSFTEDEKIALLEILSKDHBEYELIKVRRKATDSEIKKARYL
AIKLPDKNSHPKAFKAVINRAFEVLSNEKRSYIDRGDPDDROMPSGKASGF
RSGAGSGPMGSGFEDMFNSRRCGACRPPEDIPFLNAGSGPGASPGCSATPS
RSGPGFRVYTNNGSGSPPMRQOPRSQOQDAEFNNAVSQKLMLVLFITVLPML
KYLFS"

Query Match 64.0%; Score 16; DB 14; Length 41200;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgcgtgactcgtcgtt 18
|||||
Db 25021 TCGTGATCGTGTGT 25036

RESULT 43
AX008043 46340 bp DNA PAT 06-SEP-2000
LOCUS AX008043
DEFINITION Sequence 66 from Patent WO967395.
ACCESSION AX008043
VERSION AX008043.1 GI:9995728
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 46340)
AUTHORS Perlin,J.P., Rieger,F. and Alliel,P.M.
TITLE Nucleic sequence and deduced protein sequence family with human
endogenous retroviral motifs, and their uses
JOURNAL INST NAT SANTE RECH MED (FR); PERLIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALLIEL PATRICK M (FR)
FEATURES
Source
1..46340
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 16104 a 8737 c 8435 g 13064 t
ORIGIN

Query Match 64.0%; Score 16; DB 9; Length 46340;
Best Local Similarity 100.0%; Pred. No. 86;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 4 gctgactcgtcgtt 19
|||||
Db 1254 GCTGATCGTGTGT 1239

RESULT 44
AB018119
LOCUS AB018119 62927 bp DNA PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, p1 clone:MSN2.
ACCESSION AB018119 BA000015
VERSION AB018119.1 GI:3702737
KEYWORDS
SOURCE
ORGANISM Arabidopsis thaliana (strain:Columbia) DNA, clone_11b:Mitsui P1
clone:MSN2.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
1 (sites)
Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E., Kotani,H.
and Tabata,S.
Structural analysis of Arabidopsis thaliana chromosome 5. X.
Sequence features of the regions of 3,076,755 bp covered by sixty
P1 and TAC clones
PL and Res 7 (1), 31-63 (2000)
20181125
2 (bases 1 to 62927)
Nakamura,Y.
Direct Submission
Submitted (06-OCT-1998) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research, 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934)
Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934)
Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/andgraph.cgi?c=MSN2
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/Grail-1.3/).
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremi1.n1.zozi.iastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is K1F13 and the 3' clone is MUD21.
Location/Qualifiers
1..62927
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="MSN2"
/clone_11b="Mitsui P1"
/join(5223..5458,5553..5649,5719..5799,5902..6021,
6101..6313,6392..6454,6544..6721,6803..7170)
/note="EmbICAB16816.1
gene_id:MSN2.2
similar to unknown protein"
/codon_start=1
/evidence=not-experimental
/protein_id="BA97270.1"
/db_xref="GI:9843722"
/translation="MVRRKROEEDKEIEIRVKESLAKQAEEKRRLSESKGRGR

IOVDDQLAKTTTQKQGLNHSKDVVEEDVNPSSIDOKSEIDGTSNPRCLCFH
 CHRPVMEHILKKGKFIHIDCYEYRNRCYVCOQKIPVNAEGI RKFSEHPFKKCYC
 PLHDEGTAKCCSCERLEPGRNYMLGDFMLCIECMGSAMVDNENPOPLHEIFRE
 FGLFLVKEKELLVLEKQALNAKEEEKIDHYHAAYRGICNSEEDIVPSILIKGR
 MGDNDLITDIYESORVSFEVGTGIIITYGPRILITGILLAHBMHMLRNGKNL
 KLEIFEGICQALGRLMBESOTFASTDAANAAVAASSSSSTAPPAITSKSDWS
 IFEKLVFCMMQIKEDDSPYGLGFCQVYEMMSNNYNIKDTLKDIVASNAVPDST
 V"

complement(join(8609,.9043,9132,.9662))
 /note="contains similarity to unknown protein
 gblAAC17084.1
 gene_id:MSN2.3"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BA97271.1"
 /db_xref="GI:8843723"
 /translation="MSSKSLVSLVFNISKYSTRISGLTRMRPTKSGGIPPDAGDSGI
 RRRLHRRAFSEPEIVPKGNLMETRLLEISNNRIRLDEMILPPSPKSSPEEPA
 VIVEDYKLMRAEMELVSKRLRETKGNWVPSSEVRVCGEYSDPDEGNVANNMD
 AGNVIVIGKLVGKLPPELISAMAGLIPPLESLDETQEPFOLEIISDIDRADDL
 VAKELMAGLITMAQTVGFRLTFWELSDMVEPLCFVSTYFMAAGVAFPLRTSKEP
 SPEGFYKSFETKQRLIKMLDEFDIDRFTKLOKMRPMLTKSGRC"
 complement(14713,.15909)
 /note="gene_id:MSN2.4"
 /codon_start=1
 /evidence=not_experimental
 /product="At14a protein-like"
 /protein_id="BA97272.1"
 /db_xref="GI:8843724"
 /translation="MEPCGLSELNNGESSKRNRPSTLPYKEVTRDMRSYSSDLS
 YTSACKDSNLSKSPDSLSHORNTIITLSLAARETOSLNDLSIMVGLFLENNAY
 RVILSREDVWKNKDLKSIVDYEPKSTKTDPCQENVCYKRIEISOLITIRAVKOT
 EAEVDTLGDGDKKKKKTLEELNKKRKMADPDGELVTOFDSVTDQVLEELR
 KORRLDKORVNTKRTVSNVFATVAVLSVAATMSAPVCAVAVSGVSTAPI
 ETGKWFESOMKRYEKAVKROGLVLTMSERYVNNEMKNI RSDVDLRAVWSSTLE
 TVDFAVEREELAMGLAMOGIKKHVDGFTKMEVEGNAAKCSKFIALGRLVLEHI
 LGIPAN"

complement(16879,.18105)
 /note="gene_id:MSN2.5"
 /codon_start=1
 /evidence=not_experimental
 /product="At14a protein-like"
 /protein_id="BA97273.1"
 /db_xref="GI:8843725"
 /translation="MVFGLSEPLSLKSHSSNNNGTNSLAKTIOTNIRSOYSDLS
 YASACKYSSLSKSPDSLSHORNTIITLSLAARETOSLNDLSIMVGLFLENNAY
 RVILSREDVWKNKDLKSIVDYEPKSTKTDPCQENVCYKRIEISOLITIRAVKOT
 EAEVDTLGDGDKKKKKTLEELNKKRKMADPDGELVTOFDSVTDQVLEELR
 KORRLDKORVNTKRTVSNVFATVAVLSVAATMSAPVCAVAVSGVSTAPI
 ETGKWFESOMKRYEKAVKROGLVLTMSERYVNNEMKNI RSDVDLRAVWSSTLE
 TVDFAVEREELAMGLAMOGIKKHVDGFTKMEVEGNAAKCSKFIALGRLVLEHI
 LGIPAN"

complement(20757,.21976)
 /note="gene_id:MSN2.6"
 /codon_start=1
 /pseudo
 /evidence=not_experimental
 /product="At14a protein-like"
 complement(join(23245,.23292,23403,.23486,23567,.23671,
 23788,.23913,24012,.24083,24250,.24325,24468,.24509,
 24594,.24649,24801,.24929,25065,.25118,25274,.25399,
 25486,.25622,25713,.25986))
 /note="gene_id:MSN2.7"
 /codon_start=1
 /evidence=not_experimental
 /product="golicyl-di-phosphooligosaccharide-protein
 glycotransferase (oligosaccharyl transferase)-like"
 /protein_id="BA97274.1"
 /db_xref="GI:8843726"
 /translation="MVNLSRVALISVFLPLLSFSPVDNPDRLVLLDLISKS
 SHSIFNTLKSFGFDLPKLAEDSKLALORGLVNGLIIFAPSTRPGGSDSKSI
 *ADPFGSHLITADTASDLIRGIAECGCYDPEDPSAMVITHTSVSDVGDHRL
 IAADLVKSDVITLCKTIEAVLIERGVAHSLNPNLNLVLYLSASPSASNPSSKLS

SPDITGSSILSVMOVCISARRNARVIGSSVOLFEIDRLIRSGVOKAGSPNOYK
 SGNOPTETLSKWFHEHRLKGNLVHHRVDEPAIYRIKDLFESEVIEWEGK
 SMREYVAVNDVOVOVMSPYVITLSDKGLIHTSPKVDVGVQVOPVEYVKKLGYT
 TLTLSKQITPVRPYHNHYENFIFATPIYTGACTTTAGGFVFSFVILYHK"
 complement(30560,.32005)
 /note="gene_id:MSN2.8"
 /codon_start=1
 /evidence=not_experimental
 /product="UTP-glucose glucosyltransferase"
 /protein_id="BA97275.1"
 /db_xref="GI:8843727"
 /translation="MHTKPHAMSSPGMGHIVYIELCKRLSANNGFHYVLET
 DAASQKFLNIGSTGIVKLPSPDIYGLVDDHVVTKIGVIMRAAVPALRSKIAMH
 OKPALIVDLEGTALCLAKFNMLSYVETPTNARLGSIVYPNIDKDIKEBHYOR
 NPLAIPGCEVREFREDLDAVLVDEEYRDFVHGLAYKADGILVNTWEMPKSLK
 SLNPKLLGVARVPYVPIGILCRPIOSSTHVDLMIDNEOPNESVLYISRGSGCL
 SAKOITELANGLECSOORFVWVRPVPDSCGSEYVANGNGEDNTPKYLPGPVR
 TSPRGFVPMQAQAEILSRVAGELTHGNSSTYLESYVGYPAIAPLPAQONNA
 ALSDELGIVARLDDKEDI SRKITELVKRVTEKEGEMRRKVKKLKDSAMSLSI
 DGGGLAHSILVRLKECQFLREVLDNRGA"
 join(39811,.40266,40937,.41167)
 /note="gene_id:MSN2.9"
 /codon_start=1
 /evidence=not_experimental
 /product="homeodomain transcription factor-like"
 /protein_id="BA97276.1"
 /db_xref="GI:8843728"
 /translation="MDHGRIMDDOMLGSQVVPYTPONSHCIIVNIDGEESEKPV
 KRRRRKSGSSATNEEDVAIEIGMLRKRLQDEOVNMLEYSGNHEKLESKEKTA
 GELGIDROVAVFONRBAWRKKELEEVAKIKHNHVDVAGCCOLEQIITKEIOL
 SEQSETRKISEPLEEPNPNSSSSISVEANNAPTPELAPENYIIPYMLDNNTLO
 SMEWDLIV"
 join(42014,.42279,42692,.42827,42933,.43128,43201,.43322,
 43403,.43558,43628,.43969)
 /note="gene_id:MSN2.10"
 /codon_start=1
 /evidence=not_experimental
 /product="protein kinase Atn1-like protein"
 /protein_id="BA97277.1"
 /db_xref="GI:8843729"
 /translation="MRPGYRAPSMQKPTDYPTDQLHPNPFLLMSHGLKSFESD
 EDDSDSSNDQFAFTINTELLVVDKISIGDFGESSSTYVGRFVRVAVSKTFO
 PKRTAISLIPORRFOREVLILSKFPHENIVRGICTPIKMLITLEMEGNTLQPL
 LSVRPKPIDIKTISFALDIARGLMEFLNNGIITHRDKSNMILGDQKVLADPGL
 ARETEGPMFEFGCYRMMARPELFSVDTEIGKKRDKVRYVSPATYFRLINRT
 PRGKNNITVAVASKNORPSVNTLPEGVSTIQSCWAENPDRPEFKETITSLNLL
 RLSLSDTDAVSSSKANIATEDSTSLVQERVVCDGGLKMSKTKKLRKTKLMNMI
 VPLKLEFKSMK"

CDS

complement(join(44420,.44707,44833,.44982,45075,.45384,
 45454,.45950))
 /note="contains similarity to unknown protein
 emb|CAB46038.1
 gene_id:MSN2.11"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BA97278.1"
 /db_xref="GI:8843730"
 /translation="MSATLALSLNIPVSQFQRIYAGKSKSFSSGCRRLFSDSR
 RQAMASGSLPVFGDCLDDLVTCNGIDFTKRSKSGSFTTINCIPVSMRIGKRGGM
 MNKRLVCHSVVDPLEKSRALFGTSGKSVHTSPMACFSVGPAAHELSLNGQSESPPT
 TTYSLSLRLVSGCYLPPHREKATGCEBAHICDBEQAIIVADGVGMAEAVGVAGL
 FSELMSSVSVAIOEQHKGSSIDPLVLEKASHTKAKSGASACIILVLDKDLHINL
 GDSGFTVAREGTTVPQVQGNFTTYQLESNADVPSSQGVFTTIDVQSDVYVAG

Query Match 64.0%; Score 16; DB 12; Length 63927;
 Best Local Similarity 100.0%; Pred. No. 84;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 3 tgcggatctgcgtgt 18
 |||||||
 Db 55052 TgcTgcAtTcTgcTgtT 55067

```

RESULT 45
AC013956/c
LOCUS AC013956 65679 bp DNA HTG 16-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC013956
VERSION AC013956.1 GI:6437379
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 65679)
AUTHORS Adams, M. and Venter, J.C.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10211528 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source 1..65679
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 18648 a 14529 c 14360 g 18142 t
ORIGIN
Query Match 64.0%; Score 16; DB 63; Length 65679;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gttgctgactgctg 16
|||||
Db 63993 GTTGCTGATCTGCTG 63978

```

Search completed: October 9, 2001, 15:44:56
 Job time: 14682 sec

1

2

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:52:06 ; Search time 469.84 Seconds

(without alignments)
33.410 Million cell updates/sec

Title: US-09-396-196f-3

Perfect score: 25

Sequence: 1 gtgtgtgattctgtgtttgaagcg 25

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 9

Total number of hits satisfying chosen parameters: 41608

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

N.Geneseq_0601:*

1: /SIDSL/gcgdata/geneseq/geneseqn/NA1980.DAT:*

2: /SIDSL/gcgdata/geneseq/geneseqn/NA1981.DAT:*

3: /SIDSL/gcgdata/geneseq/geneseqn/NA1982.DAT:*

4: /SIDSL/gcgdata/geneseq/geneseqn/NA1983.DAT:*

5: /SIDSL/gcgdata/geneseq/geneseqn/NA1984.DAT:*

6: /SIDSL/gcgdata/geneseq/geneseqn/NA1985.DAT:*

7: /SIDSL/gcgdata/geneseq/geneseqn/NA1986.DAT:*

8: /SIDSL/gcgdata/geneseq/geneseqn/NA1987.DAT:*

9: /SIDSL/gcgdata/geneseq/geneseqn/NA1988.DAT:*

10: /SIDSL/gcgdata/geneseq/geneseqn/NA1989.DAT:*

11: /SIDSL/gcgdata/geneseq/geneseqn/NA1990.DAT:*

12: /SIDSL/gcgdata/geneseq/geneseqn/NA1991.DAT:*

13: /SIDSL/gcgdata/geneseq/geneseqn/NA1992.DAT:*

14: /SIDSL/gcgdata/geneseq/geneseqn/NA1993.DAT:*

15: /SIDSL/gcgdata/geneseq/geneseqn/NA1994.DAT:*

16: /SIDSL/gcgdata/geneseq/geneseqn/NA1995.DAT:*

17: /SIDSL/gcgdata/geneseq/geneseqn/NA1996.DAT:*

18: /SIDSL/gcgdata/geneseq/geneseqn/NA1997.DAT:*

19: /SIDSL/gcgdata/geneseq/geneseqn/NA1998.DAT:*

20: /SIDSL/gcgdata/geneseq/geneseqn/NA1999.DAT:*

21: /SIDSL/gcgdata/geneseq/geneseqn/NA2000.DAT:*

22: /SIDSL/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	20	AA01303
2	25	100.0	1084	10	AA01329
3	25	100.0	1121	7	AA060496
4	25	100.0	5872	15	AA062386
5	16	64.0	5000	21	AA236139
6	15	60.0	1683	16	AA214850
7	15	60.0	1683	16	AA208746
8	15	60.0	1689	20	AA230324
9	15	60.0	2022	21	AA241132
10	15	60.0	2034	20	AA204506
11	15	60.0	3079	20	AA230325

12	15	60.0	6703	19	AA049536
13	14	56.0	275	20	AA212718
14	14	56.0	336	22	AA055487
15	14	56.0	379	20	AA087319
16	14	56.0	383	21	AA026529
17	14	56.0	384	18	AA010888
18	14	56.0	612	21	AA039862
19	14	56.0	619	21	AA044124
20	14	56.0	707	21	AA014018
21	14	56.0	735	16	AA025052
22	14	56.0	753	20	AA020347
23	14	56.0	769	21	AA015012
24	14	56.0	891	20	AA026233
25	14	56.0	1044	21	AA082834
26	14	56.0	1155	20	AA020984
27	14	56.0	1155	22	AA099433
28	14	56.0	1173	20	AA020358
29	14	56.0	1220	21	AA044861
30	14	56.0	1367	21	AA081673
31	14	56.0	1493	18	AA010889
32	14	56.0	1676	22	AA030329
33	14	56.0	1848	18	AA096346
34	14	56.0	1998	20	AA020367
35	14	56.0	2080	20	AA030162
36	14	56.0	2087	20	AA020343
37	14	56.0	2166	20	AA020379
38	14	56.0	2405	22	AA027417
39	14	56.0	2574	20	AA233977
40	14	56.0	2574	21	AA078481
41	14	56.0	2574	21	AA058239
42	14	56.0	2669	20	AA225020
43	14	56.0	2675	21	AA098021
44	14	56.0	3243	11	AA005686
45	14	56.0	3243	20	AA232338

ALIGNMENTS

RESULT 1

AA01303

ID AA01303 standard; DNA; 1041 BP.

AC AA01303;

XX

DT 12-APR-1999 (first entry)

XX

DE E. coli biotin synthetase (Biot) coding sequence.

XX

KW DAP aminotransferase; diaminopelargonic acid; transgenic plant;

KW biotin synthase; biotin production; vitamin H; Biot; ss.

XX

OS Escherichia coli.

XX

PN US5869719-A.

XX

PD 09-FEB-1999.

XX

PF 30-APR-1997; 97US-0846338.

XX

PR 30-APR-1997; 97US-0846338.

PR 08-MAR-1995; 95US-0401068.

XX

PA (NOVS) NOVARTIS FINANCE CORP.

PI Patton DA;

PI

DR WPI: 1999-152902/13.

DR P-PSDB; AA073906.

XX

PT Transgenic plants with high biotin levels - transformed with DNA

PT encoding diaminopelargonic acid amino-transferase or biotin

PT synthase

XX Example 2; Column 37-40; 34pp; English.

CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC be used in the transgenic plant of the invention. The transgenic plant,
CC plant cell or plant tissue is transformed with a chimeric gene encoding
CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
CC produces more biotin than a non-transgenic plant, cell or tissue. The
CC plant is used as an improved dietary source of biotin (vitamin H) for
CC humans or animals.

SQ Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtgtcgtgactgctgtttgaagcg 25
|||
Db 57 gtgtcgtgactgctgtttgaagcg 81

RESULT 2

ID AAN91329 standard; DNA; 1084 BP.

XX AAN91329;

DT 15-FEB-1990 (first entry)

DE E.coli Bio B gene.

KW E.coli; Bio B gene; biotin.

OS Escherichia coli.

FT Key Location/Qualifiers
CDS 24..1064
/*tag=a

PN GB2216530-A.

XX 11-OCT-1989.

PF 17-MAR-1989; 89GB-0006210.

PR 22-MAR-1988; 88GB-0006804.

PR 17-MAR-1989; 89GB-0006210.

PA (UKAG-) UK MIN. AGRIC. FISH.

PI Pearson BM, McKee RA;

DR WPI; 1989-295085/41. P-PSDB P91392

PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes
PT - derived from E.coli and capable of replication and expression in other
PT microorganisms, esp. yeast.

PS Table 3; page 33-4; 52pp; English.

CC The gene can be used in a plasmid for expression of enzymes of the biotin
CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
CC Lactobacillus. Insertion of bio B improves biotin yields in
CC microorganisms which export biotin, or enables growth in media contg.
CC little or no biotin of organisms unable to synthesise biotin for their
CC own use.

SQ Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtgtcgtgactgctgtttgaagcg 25
|||
Db 80 gtgtcgtgactgctgtttgaagcg 104

RESULT 3

ID AAN60496 standard; DNA; 1121 BP.

XX AAN60496;

DT 17-OCT-1991 (first entry)

DE Sequence encoding biotin synthesising enzyme.

KW Biotin synthetic enzyme; E.coli; desthiobiotin; ds.

FT Key Location/Qualifiers
CDS 42..1082
/*tag= a

PN JF61149091-A.

XX 07-JUL-1986.

PF 24-DEC-1984; 84JP-0272605.

PR 24-DEC-1984; 84JP-0272605.

PA (NIPS) NIPPON SODA KK.

XX WPI; 1986-216622/33.

DR P-PSDB; AAP60536.

PT Double stranded DNA encoding biotin synthesising enzyme -

PT comprises transformed mutant E.coli strain contg. cyclic doubled

PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.

PS Disclosure; Page 534; 23pp; Japanese.

XX The sequence may be expressed by a transformed E.coli host, cultured

CC in a medium containing desthiobiotin.

SQ Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtgtcgtgactgctgtttgaagcg 25
|||
Db 98 gtgtcgtgactgctgtttgaagcg 122

RESULT 4

ID AAO62386 standard; DNA; 5872 BP.

XX AAO62386;

DT 16-NOV-1994 (first entry)

DE Biotin-biosynthesis genes contg. plasmid pB030A-15/9.

KW Biotin; expression; enterobacteria; vitamin H; synthesis;

KW plasmid; pB030A-15/9; bioB; bioC; bioD; bioA;

KW promoter ptac; biotin synthase; KAPA synthase;

KW 8-amino-7-oxononanoate synthase; pimeioyl-CoA; DFB synthase;

```

KW dethiobiotin synthase; DAPA synthase;
KM S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
XX Seborrhoea; dermatitis; ds.
OS Escherichia coli DSM498.
FH Key Location/Qualifiers
FH promoter 1..96
FH /tag= a
FH /function= "promoter ptac"
FH /evidence= EXPERIMENTAL
FH -35_signal 23..28
FH /tag= b
FH /standard_name= "promoter ptac"
FH -10_signal 45..50
FH /tag= c
FH /evidence= EXPERIMENTAL
FH /standard_name= "promoter ptac"
FH 105..109
FH /tag= d
FH /evidence= EXPERIMENTAL
FH /standard_name= "bioB RBS no. 9"
FH 117..1157
FH /tag= e
FH /product= "biotin synthase"
FH /evidence= EXPERIMENTAL
FH /gene= "bioB"
FH /number= 1
FH 1141..1146
FH /tag= f
FH /standard_name= "bioF RBS"
FH 1154..2311
FH /tag= g
FH /EC_number= 2.3.1.47
FH /product= "KAPA synthase"
FH /evidence= EXPERIMENTAL
FH /gene= "bioF"
FH /number= 2
FH /standard_name= "8-amino-7-oxononanoate synthase"
FH 2284..2288
FH /tag= h
FH /standard_name= "bioC RBS"
FH 2295..3050
FH /tag= i
FH /function= "involved in pimeloyl-CoA synthesis"
FH /product= "protein"
FH /gene= "bioC"
FH /number= 3
FH 3030..3033
FH /tag= j
FH /standard_name= "bioD RBS"
FH 3043..3753
FH /tag= k
FH /EC_number= 6.3.3.3
FH /product= "DTB synthase"
FH /evidence= EXPERIMENTAL
FH /gene= "bioD15"
FH /number= 4
FH /standard_name= "dethiobiotin synthase"
FH 3712..3750
FH /tag= l
FH /note= "bioD15 substitution"
FH 3742..3746
FH /tag= m
FH /standard_name= "bioA RBS"
FH 3750..5039
FH /tag= n
FH /EC_number= 2.6.1.62
FH /product= "DAPA synthase"
FH /evidence= EXPERIMENTAL
FH /gene= "bioA"
FH /number= 5
FH /standard_name= "S-adenosyl-L-methionine: 8-amino-

```

```

FH RBS 5088..5093
FH /tag= o
FH /standard_name= "ORF1 RBS"
FH 5098..5574
FH CDS
FH /tag= p
FH /function= "unknown, involved in biotin synthase"
FH /product= "protein"
FH /evidence= EXPERIMENTAL
FH /gene= "ORF1"
FH /number= 6
FH terminator 5583..5644
FH /tag= q
FH /standard_name= "rho-independent transcriptional
FH stem_loop 5583..5605
FH /tag= r
FH W09408023-A.
FH 14-APR-1994.
FH PD 14-APR-1994.
FH XX 01-OCT-1993; 93WO-EP02688.
FH PF 01-OCT-1993; 93CH-0003124.
FH PR 02-OCT-1992; 93CH-0002134.
FH PR 15-JUL-1993; 93CH-0002134.
FH XX (LONZ ) LONZA AG.
FH PA Birch O, Brass J, Fuhrmann M, Shaw N;
FH PI WPI: 1994-135587/16.
FH XX P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
FH DR Biotechnological biotin prodn. using enterobacterial biotin-gene
FH XX - providing vitamin H in high yield
FH PT Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
FH PS The sequence is derived from plasmid pB030A-15/9 contg. the
FH CC bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
FH CC of biotin, arranged in a transcription unit. Microorganisms
FH CC contg. these DNA fragments or plasmids may be used in the prodn.
FH CC of biotin. Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
FH CC loss of appetite and tiredness.
FH XX Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;
SQ
Query Match 100.0%; Score 25; DB 15; Length 5872;
Best Local Similarity 100.0%; Pred. NO. 0.00013;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 gtgtcgtgacgtcgtcgttgaagcg 25
DB 173 gtgtcgtgacgtcgtcgttgaagcg 197
RESULT 5
AAZ36139
ID AAZ36139 standard; DNA: 5000 BP.
AC AAZ36139;
DE 11-FEB-2000 (first entry)
Nucleotide sequence of the DDM1 gene which regulates DNA methylation.
XX DDM1 gene; chromosome 5; zinc-finger protein gene; SM12/SMF2 family;
KW glutamic acid tRNA gene; genomic cytosine methylation; gene silencing;
KW DNA methylation; chromatin structure; DNA hypomethylation;
XX DDM1 gene allele; inbreeding depression; ss.
XX

```

OS	Arabidopsis thaliana.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	535..4354
FT		/*tag= a
FT	exon	/note= "contains introns"
FT		535..5566
FT		/*tag= b
FT	intron	/number= 1
FT		567..771
FT		/*tag= c
FT	exon	/number= 1
FT		772..850
FT		/*tag= d
FT	allele	/number= 2
FT		replace(785, (N82))
FT		/*tag= e
FT		/note= "Site of the ddm1-5 (som8) mutation; this base is replaced with 82 unspecified bases"
FT	intron	851..985
FT		/*tag= f
FT	exon	/number= 2
FT		986..1252
FT		/*tag= g
FT	intron	/number= 3
FT		1253..1353
FT		/*tag= h
FT	exon	/number= 3
FT		1354..1440
FT		/*tag= i
FT	intron	/number= 4
FT		1441..1548
FT		/*tag= j
FT	exon	/number= 4
FT		1549..1895
FT		/*tag= k
FT	intron	/number= 5
FT		1896..1975
FT		/*tag= l
FT	exon	/number= 5
FT		1976..2165
FT		/*tag= m
FT	intron	/number= 6
FT		2164..2250
FT		/*tag= n
FT	exon	/number= 6
FT		2251..2426
FT		/*tag= o
FT	allele	/number= 7
FT		replace(2384..2385)
FT		/*tag= p
FT		/note= "Site of the ddm1-6 mutation; either the G at 2384 or 2385 is deleted"
FT	intron	2427..2558
FT		/*tag= q
FT	exon	/number= 7
FT		2559..2625
FT		/*tag= r
FT	intron	/number= 8
FT		2626..2702
FT		/*tag= s
FT	exon	/number= 8
FT		2703..2892
FT		/*tag= t
FT	intron	/number= 9
FT		2893..2974
FT		/*tag= u
FT	exon	/number= 9
FT		2975..3070
FT		/*tag= v
FT	intron	/number= 10
FT		3071..3147
FT		/*tag= w

FT	/number= 10
FT	3148..3242
FT	/*tag= x
FT	/number= 11
FT	3186
FT	/*tag= y
FT	/note= "alternative splice donor site used in ddm1-2"
FT	3243..3316
FT	/*tag= z
FT	/number= 11
FT	replace(3243, A)
FT	/*tag= aa
FT	/note= "site of the ddm1-2 mutation"
FT	3317..3436
FT	/*tag= ab
FT	/number= 12
FT	replace(3337, A)
FT	/*tag= ac
FT	/note= "site of the ddm1-7 (som5) mutation"
FT	3437..3539
FT	/*tag= ad
FT	/number= 12
FT	3540..3659
FT	/*tag= ae
FT	/number= 13
FT	3660..3744
FT	/*tag= af
FT	/number= 13
FT	3745..3843
FT	/*tag= ag
FT	/number= 14
FT	3844..3933
FT	/*tag= ah
FT	/number= 14
FT	3934..4038
FT	/*tag= ai
FT	/number= 15
FT	4039..4129
FT	/*tag= aj
FT	/number= 15
FT	4130..4354
FT	/*tag= ak
FT	/number= 16
FT	4355..4826
FT	/*tag= al
FT	/note= "complement of predicted tRNA-glu"
FN	
FX	
FX	W09955891-A1.
PX	
PD	04-NOV-1999.
XX	
PE	29-APR-1999; 99WO-USO9268.
XX	
PR	24-JUN-1998; 98US-0104070.
PA	(UNIW) UNIV WASHINGTON.
XX	
PI	Richards EJ, Jeddeloh JA;
XX	
DR	WPI; 2000-013439/01.
DR	P-PSDB; AAY43847.
PT	Arabidopsis thaliana DDM1 gene alters methylation state of DNA, useful
PT	for reducing gene silencing or inbreeding depression -
XX	
PS	Claim 6; Page 36-38; 59pp; English.
XX	
CC	The present sequence represents the DDM1 gene of Arabidopsis thaliana.
CC	The gene is located on chromosome 5, lower arm, where it is flanked on
CC	the centromeric side, within 20 kb, by a zinc-finger protein gene and
CC	on the telomeric side, within 1 kb, by a glutamic acid tRNA gene. The
CC	DDM1 protein is a member of the SWI2/SNF2 family of polypeptides. The
CC	DDM1 protein is important for the maintenance of proper genomic cytosine

CC methylation, and its function appears to be necessary to maintain gene
CC silencing. The protein appears to function in the DNA methylation system
CC by affecting chromatin structure. Disruption of the gene is associated
CC with DNA hypomethylation. Four DDM1 gene alleles have been identified,
CC all of which are expected to destroy or severely reduce gene function.
CC The DDM1 nucleic acid can be used to create transgenic plants with
CC altered levels of DNA methylation. The mutant or engineered plants are
CC expected to be viable and display valuable characteristic such as
CC inbreeding depression or a reduction in gene silencing.

SQ Sequence 5000 BP; 1386 A; 909 C; 1029 G; 1676 T; 0 other;

Query Match 64.0%; Score 16; DB 21; Length 5000;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgctgagctctgctgt 18
|||||
Db 1132 tgctgagctctgctgt 1147

RESULT 6
AAT14850/C
ID AAT14850 standard; DNA; 1683 BP.

AC AAT14850;

DT 04-JUL-1996 (first entry)

DE Maize U14.4 DNA sequence.

XX snORNA; small nucleolar RNA; U14; non-translated sequence; conserved;
KW Box C; Box D; processing; pre-ribosomal RNA; ribosome formation;
KW normal growth phenotype; yeast; plant; maize; potato; 18S rRNA; primer;
XX PCR; polymerase chain reaction; amplification; ss.

OS Zea mays.

PH Key Location/Qualifiers
FT misc_feature 379..385
/*tag= a
/note= "nucleotides not identified in sequence"

FT repeat_unit 1212..1217
/*tag= b
/note= "inverted repeat"

FT misc_RNA 1218..1329
/*tag= c
/note= "MU14.4"

FT misc_feature 1220..1225
/*tag= d
/note= "Box C"

FT misc_feature 1324..1329
/*tag= e
/note= "Box D"

FT repeat_unit 1330..1335
/*tag= f
/note= "inverted repeat"

PN WO9530748-A2.

PD 16-NOV-1995.

XX 04-MAY-1995; 95WO-EP01694.

XX 09-DEC-1994; 94EP-0119487.

XX 04-MAY-1994; 94WO-EP01409.

XX (GENE-) GENE SHEARS PTY LTD.

PI Brown JWS, Leader DJ, Waugh R;

XX MPI; 1995-404115/51.

XX plant U14 nucleic acid sequences and derivatives - regulate
PT ribosomal RNA production and accumulation in plants
XX Example; Fig 7A-D; 57pp; English.

XX The present sequence shows the maize single U14 sequence U14.4,
CC including flanking intergenic sequences and inverted repeats, as obtd.
CC by sequencing of fragments of isolated genomic clones. Inverted repeats
CC and Box C and D elements are shown above (see AAT08751-52). U14 genes
CC are small nucleolar RNAs present in the genome as non-translated
CC sequences. snORNAs are involved in processing of pre-ribosomal RNA
CC transcripts and ribosome formation. U14snORNA is required for a normal
CC growth phenotype and for processing of pre-rRNA transcripts in yeast.
CC Plant U14 sequences have been identified and characterised in the present
CC invention, and may be intron encoded.

SQ Sequence 1683 BP; 383 A; 387 C; 358 G; 548 T; 7 other;

Query Match 60.0%; Score 15; DB 16; Length 1683;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtgtgctgagctctgt 15
|||||
Db 601 GTTGCTGATCTGCT 587

RESULT 7
AAT08746/C
ID AAT08746 standard; DNA; 1683 BP.

AC AAT08746;

DT 04-JUL-1996 (first entry)

DE Maize U14.4 DNA sequence.

XX snORNA; small nucleolar RNA; U14; non-translated sequence; conserved;
KW Box C; Box D; processing; pre-ribosomal RNA; ribosome formation;
KW normal growth phenotype; yeast; plant; maize; potato; 18S rRNA; primer;
XX PCR; polymerase chain reaction; amplification; ss.

OS Zea mays.

PH Key Location/Qualifiers
FT misc_feature 379..385
/*tag= a
/note= "nucleotides not identified in sequence"

FT repeat_unit 1212..1217
/*tag= b
/note= "inverted repeat"

FT misc_RNA 1218..1329
/*tag= c
/note= "MU14.4"

FT misc_feature 1220..1225
/*tag= d
/note= "Box C"

FT misc_feature 1324..1329
/*tag= e
/note= "Box D"

FT repeat_unit 1330..1335
/*tag= f
/note= "inverted repeat"

PN WO9530747-A1.

PD 16-NOV-1995.

XX 04-MAY-1994; 94WO-EP01409.

XX 04-MAY-1994; 94WO-EP01409.

XX (GENE-) GENE SHEARS PTY LTD.
 XX Brown JWS, Leader DJ, Waugh R;
 XX WPI: 1995-404114/51.
 XX
 PT Nucleic acid contg. plant U14 sequences and their derivs. - for
 PT regulating prodn. of ribosomal RNA, also as stabilisers for
 PT heterologous RNA
 XX
 PS Example: Fig 7A-D; 57pp; English.
 XX
 CC The present sequence shows the maize single U14 sequence U14.4,
 CC including flanking intergenic sequences and inverted repeats, as obtd.
 CC by sequencing of fragments of isolated genomic clones. Inverted repeats
 CC and Box C and D elements are shown above (see AAT08729-30). U14 genes
 CC are small nucleolar RNAs present in the genome as non-translated
 CC sequences. SnORNAs are involved in processing of pre-ribosomal RNA
 CC transcripts and ribosome formation. U14snRNA is required for a normal
 CC growth phenotype and for processing of pre-rRNA transcripts in yeast.
 CC Plant U14 sequences have been identified and characterised in the present
 CC invention, and may be intron encoded.
 XX
 SO Sequence 1683 BP; 383 A; 387 C; 358 G; 548 T; 7 other;

Query Match 60.0%; Score 15; DB 16; Length 1683;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtgtcgtgactgtct 15
 |||||
 Db 601 GTTGCTGATCTGCT 587

RESULT 8
 AA223024/C
 ID AA223024 standard; cDNA; 1689 BP.
 XX
 AC AA223024;
 XX
 DT 17-JAN-2000 (first entry)
 XX
 DE Rat kd312 polypeptide encoding cDNA.
 XX
 KM kd312 polypeptide; cancer; immunodeficiency disorder; AIDS; stroke;
 KM heart attack; head trauma; neurodegenerative disease; rat;
 KM Parkinson's disease; Alzheimer's disease; ss.
 XX
 OS Rattus sp.
 XX
 PN WO9950288-A2.
 XX
 PD 07-OCT-1999.
 XX
 PE 30-MAR-1999; 99WO-US06993.
 XX
 PR 31-MAR-1998; 98US-0053374.
 XX
 PA (AMGE-) AMGEN INC.
 XX
 PI Yen K;
 XX
 DR WPI: 1999-601322/51.
 DR P-PSDB; AAY42694.
 XX
 XX kd312 polypeptides useful for treating diseases and disorders
 PT associated with alterations in cell proliferation and cell death
 PT
 PS Claim 2; Fig 9; .95pp; English.
 XX
 CC The invention provides nucleic acid molecules encoding human and rat

CC kd312 polypeptides. The kd312 polypeptides can be expressed by standard
 CC recombinant methodology. The kd312 sequences, and the antibodies against
 CC the proteins may be used to treat or diagnose the presence or progression
 CC of various diseases, e.g. cancer, immunodeficiency disorders (e.g. AIDS),
 CC stroke, heart attack, head trauma, and neurodegenerative diseases (e.g.
 CC Parkinson's disease and Alzheimer's disease). The present sequence
 CC represents the rat kd312 cDNA sequence.
 XX
 SO Sequence 1689 BP; 417 A; 462 C; 434 G; 376 T; 0 other;

Query Match 60.0%; Score 15; DB 20; Length 1689;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 gacactgctgttga 21
 |||||
 Db 495 GCATCTGCTGTTTGA 481

RESULT 9
 AAC44132/C
 ID AAC44132 standard; DNA; 2022 BP.
 XX
 AC AAC44132;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Zea mays DNA fragment SEQ ID NO: 41750.
 XX
 KM Hybridisation assay; genetic mapping; gene expression control;
 KM protein identification; signal transduction pathway; metabolic;
 KM pathway; promoter; termination sequence; corn; ss.
 XX
 OS Zea mays subsp. mays.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 XX 25-FEB-2000; 2000EP-0301439.
 XX
 PE 25-FEB-1999; 99US-0121825.
 XX
 PR 05-MAR-1999; 99US-0123180.
 XX
 PR 09-MAR-1999; 99US-0123548.
 XX
 PR 23-MAR-1999; 99US-0125788.
 XX
 PR 25-MAR-1999; 99US-0126264.
 XX
 PR 29-MAR-1999; 99US-0126785.
 XX
 PR 01-APR-1999; 99US-0127462.
 XX
 PR 06-APR-1999; 99US-0128234.
 XX
 PR 08-APR-1999; 99US-0128714.
 XX
 PR 16-APR-1999; 99US-0129845.
 XX
 PR 19-APR-1999; 99US-0130077.
 XX
 PR 21-APR-1999; 99US-0130449.
 XX
 PR 23-APR-1999; 99US-0130510.
 XX
 PR 28-APR-1999; 99US-0130891.
 XX
 PR 30-APR-1999; 99US-0131449.
 XX
 PR 30-APR-1999; 99US-0132048.
 XX
 PR 30-APR-1999; 99US-0132407.
 XX
 PR 04-MAY-1999; 99US-0132484.
 XX
 PR 05-MAY-1999; 99US-0132485.
 XX
 PR 06-MAY-1999; 99US-0132486.
 XX
 PR 06-MAY-1999; 99US-0132487.
 XX
 PR 07-MAY-1999; 99US-0132863.
 XX
 PR 11-MAY-1999; 99US-0134256.
 XX
 PR 14-MAY-1999; 99US-0134218.
 XX
 PR 14-MAY-1999; 99US-0134219.
 XX
 PR 14-MAY-1999; 99US-0134221.
 XX
 PR 14-MAY-1999; 99US-0134370.
 XX
 PR 18-MAY-1999; 99US-0134768.
 XX
 PR 19-MAY-1999; 99US-0134941.
 XX
 PR 20-MAY-1999; 99US-0135124.
 XX
 PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142877.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149922.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153756.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159328.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.

PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 60.0%; Score 15; DB 21; Length 2022;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tctgctgttgaagc 24
|||||
DB 1268 TCTGCTGTGAGC 1254

RESULT 10
AA04606/C
ID AA04606 standard; DNA; 2034 BP.
XX
AC AA04606;
XX
DT 07-APR-1999 (first entry)

DE Maize genomic clone Mzu14.4 snor cluster sequence.
XX
XX Small nucleolar RNA gene; snor; snor1.4; maize; promoter;
KM RNA production; RNA stability; small non-translated RNA;
KW intergenic region; U49.1; Snor2.4; U14.4; Snor3.4; ds.
XX
OS Zea sp.

XX
FH Key Location/Qualifiers
FT 1..890
FT misc_feature /tag= a
FT /note= "upstream region having promoter activity"
FT 891..983
FT /tag= b
FT /note= "Snor1.4 coding sequence; no protein given in
FT specification"
FT 984..1100
FT /tag= c
FT /note= "intergenic region"
FT 1101..1294
FT /tag= d
FT /note= "U49.1 coding sequence; no protein given in
FT specification"
FT 1295..1361
FT /tag= e
FT /note= "intergenic region"
FT 1362..1512
FT /tag= f
FT /note= "Snor2.4 coding sequence; no protein given in
FT specification"
FT 1513..1587
FT /tag= g
FT /note= "intergenic region"
FT 1588..1711
FT /tag= h
FT /note= "U14.4 coding sequence; no protein given in
FT specification"
FT 1712..1743
FT /tag= i
FT /note= "intergenic region"
FT 1744..1873
FT /tag= j
FT /note= "Snor3.1 coding sequence; no protein given in
FT specification"
FT 1874..2034
FT /tag= k
FT /note= "3' flanking sequence"

XX
XX EP887405-A1.
XX
XX 30-DEC-1998.
XX

PF 25-JUN-1997; 97EP-0401480.
XX
XX 25-JUN-1997; 97EP-0401480.
PR
XX
PA (GENE-) GENE SHEARS PTY LTD.
XX
PI Brown JMS, Leader DJ;

DR WPI; 1999-047872/05.
XX
XX
PT Producing stabilised RNA molecules - using a DNA precursor
PT comprising plant small nucleolar RNA stabilising sequences
XX
XX
PS Claim 46; Fig 18a-f; 98pp; English.

XX
XX The present sequence represents the full sequence of the small
CC nucleolar RNA (snorRNA) cluster of maize genomic clone Mzu14.4. The
CC snor genes in the cluster are Snor1.4, U49.1, Snor2.4, U14.4, and
CC Snor3.1. Plant snor genes enable production of multiple stabilised
CC RNA from a single promoter in a splicing-independent manner,
CC and allow the production of RNA from intron or non-intron sequences.
CC The present sequence is used in the method of the invention. The
CC specification describes a method for producing small non-translated
CC RNA molecules, which comprises introducing a precursor DNA containing
CC one or a cluster of coding sequences into a plant or yeast cell. Each
CC coding sequence comprises a non-translated RNA sequence to be stabilised,
CC which is capable of functionally interacting with a cellular component
CC and stabilising sequences comprising at least one protein binding site.
CC The stabilised RNA units are transcribed by the cell's endogenous
CC processing machinery. The process is useful for qualitatively and
CC quantitatively enhancing the production and stability of RNA molecules,
CC especially non-translated RNA e.g. ribozymes and antisense RNA. The
CC nucleic acid molecules are useful as probes or primers for identification
CC of snorRNA sequences and intergenic regions.
XX

SQ Sequence 2034 BP; 449 A; 504 C; 435 G; 646 T; 0 other;

Query Match 60.0%; Score 15; DB 20; Length 2034;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgctgagctgct 15
|||||
DB 971 GTGCTGAGCTGCT 957

RESULT 11
AA23025/C
ID AA23025 standard; DNA; 3079 BP.
XX

AC AA23025;

DT 17-JAN-2000 (first entry)

DE Rat Kd312 genomic DNA sequence.

XX
XX Kd312 polypeptide; cancer; immunodeficiency disorder; AIDS; stroke;
KM heart attack; head trauma; neurodegenerative disease; rat;
KW Parkinson's disease; Alzheimer's disease; ss.
XX

OS Rattus sp.

PN WO950288-A2.

XX
XX 07-OCT-1999.

XX
XX 30-MAR-1999; 99WO-US06993.

XX
XX 31-MAR-1998; 98US-0053374.

XX
XX (AMGE-) AMGEN INC.
XX

PI Yen K;
 XX
 DR WPI: 1999-601322/51.
 DR P-PSDB: AAV42694.
 XX
 PT k3312 polypeptides useful for treating diseases and disorders
 XX associated with alterations in cell proliferation and cell death
 PS Claim 2: Fig 7: 85pp; English.
 XX
 CC The invention provides nucleic acid molecules encoding human and rat
 CC k3312 polypeptides. The k3312 polypeptides can be expressed by standard
 CC recombinant methodology. The k3312 sequences, and the antibodies against
 CC the proteins may be used to treat or diagnose the presence or progression
 CC of various diseases, e.g. cancer, immunodeficiency disorders (e.g. AIDS),
 CC stroke, heart attack, head trauma, and neurodegenerative diseases (e.g.
 CC Parkinson's disease and Alzheimer's disease). The present sequence
 CC represents the rat k3312 genomic DNA sequence.
 XX
 SQ Sequence 3079 BP; 708 A; 843 C; 829 G; 699 T; 0 other;

Query Match 60.0%; Score 15; DB 20; Length 3079;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ggatctgctgttga 21
 |||||||
 DB 803 GGATCTGCTGTGCA 789

RESULT 12
 AAV49536
 ID AAV49536 standard; DNA: 6703 BP.
 XX
 AC AAV49536;
 XX
 DT 16-NOV-1998 (first entry)
 XX
 DE Adenylate cyclase gene.
 XX
 KM ds: adenylate cyclase; recombinant microorganism; chemical production;
 KM riboflavin; carotenoid.
 XX
 OS Ashbya gossypii.
 XX
 FH Key Location/Qualifiers
 FT 1..670
 FT /*lag= a
 FT CDS 671..6295
 FT /*lag= b
 FT /product= "Adenylate cyclase"
 FT 3'UTR 6296..6703
 FT /*lag= c
 XX
 PN WO9829538-A2.
 XX
 PD 09-JUL-1998.
 XX
 PE 29-DEC-1997; 97WO-EP07309.
 XX
 PR 31-DEC-1996; 97CH-0000016.
 XX
 PA (BADI) BASF AG.
 PI Althoefer H, Altmann-Joehl R, Philippsen P, Seubberger H;
 XX WPI: 1998-388120/33.
 DR P-PSDB: AAM64518.
 XX
 PT New gene for adenylate cyclase from Ashbya gossypii - useful for
 PT generating recombinant microorganisms with alteration in gene of
 PT cAMP-dependent signalling pathway for increasing production of fine

PT chemicals
 XX
 PS Claim 1: Page 10-20; 33pp; German.
 XX
 CC The adenylate cyclase (AC) gene is used to generate recombinant
 CC microorganisms for production of chemicals, particularly riboflavin but
 CC also others useful in human or animal nutrition, cosmetics and medicines,
 CC and also carotenoids for use as natural colourings. The modified
 CC microorganisms produce chemicals at higher levels than wild-type
 CC organisms. Particularly, for riboflavin, cells with a disrupted AC gene
 CC are grown on glucose-containing medium.
 XX
 SQ Sequence 6703 BP; 1954 A; 1527 C; 1541 G; 1681 T; 0 other;

Query Match 60.0%; Score 15; DB 19; Length 6703;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgctgattctgt 17
 |||||||
 DB 6234 tgctgattctgt 6248

RESULT 13
 AA212718
 ID AA212718 standard; CDNA: 275 BP.
 XX
 AC AA212718;
 XX
 DT 12-OCT-1999 (first entry)
 XX
 DE Human gene expression product CDNA sequence SEQ ID NO:187.
 XX
 KM Human; gene: gene expression product; diagnosis; therapy; probe;
 KM detection; mapping; tissue typing; profiling; forensic; cancer;
 KM genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9938972-A2.
 XX
 PD 05-AUG-1999.
 XX
 PF 28-JAN-1999; 99WO-US01619.
 XX
 PR 03-APR-1998; 98US-0080666.
 PR 28-JAN-1998; 98US-0072910.
 PR 24-FEB-1998; 98US-0075954.
 PR 31-MAR-1998; 98US-0080114.
 PR 03-APR-1998; 98US-0080515.
 XX
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX
 PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
 PI Escobedo J, Garcia PD, Garcia V, Glese K, Innis MA;
 PI Jamson WL, Kassam A, Kennedy GC, Kita D, Labat I;
 PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
 PI Stache-Crain B, Sudduth-Klingner J, Williams LT;
 XX WPI: 1999-494092/41.
 DR
 XX
 PT Novel human genes and their expression products which are
 PT differentially expressed in different cell types
 PS Claim 1: Page 705; 2479pp; English.
 XX
 CC The present invention describes a library of human polynucleotides
 CC comprising the sequences given in AA212532 to AA217779. Also described is
 CC a method of detecting differentially expressed genes correlated with the
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample from a cell

CC suspected of being cancerous, where the gene product is encoded by one
 CC of the 5248 polynucleotide sequences given in AA212532 to AA217779. The
 CC polynucleotides can be used as a source of primers and probes, which can
 CC be used for a variety of purpose, e.g. detection of expression levels,
 CC mapping, tissue typing or profiling, forensics, genetic analysis and
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
 CC can be used for raising antibodies for experimental, diagnostic and
 CC therapeutic purposes. The polynucleotides may also be used to construct
 CC arrays for diagnostics (which may be used to determine function of an
 CC encoded protein); and to detect differences in expression levels between
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
 CC identify a genetic predisposition or susceptibility to a disease such as
 CC cancer). The polynucleotides of the invention are especially used in the
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
 CC and lung cancer. The polynucleotides can also be used to screen for
 CC peptide analogues and antagonists.

SQ Sequence 275 BP; 65 A; 44 C; 48 G; 66 T; 52 other;

Query Match 56.0%; Score 14; DB 20; Length 275;
 Best Local Similarity 100.0%; Pred. No. 80;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ttgtgtgattctgt 15
 |||||
 Db 169 ttgtgtgattctgt 182

RESULT 14
 AAF65487/C
 ID AAF65487 standard; cDNA; 336 BP.

XX AAF65487;

XX 09-APR-2001 (first entry)

DE Novel human polynucleotide, SEQ ID NO: 1243.

XX Human; cytostatic; gene therapy; colon cancer; prostate cancer;
 KM breast cancer; lung cancer; cancer detection; ss.

XX Homo sapiens.

XX WO200102568-A2.

XX 11-JAN-2001.

XX 30-JUN-2000: 2000WO-US18374.

XX 02-JUL-1999; 99US-0142310.

XX 02-JUL-1999; 99US-0142311.

PA (CHIR) CHIRON CORP.

XX (HYSE) HYSE INC.

PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
 PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
 PI Kirkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
 PI Kita D, Garcia V, Jones LW, Strache-Crain B;

XX WPI: 2001-091805/10.

PT Library of polynucleotides for diagnosing a cancerous state of a
 PT mammalian cell and detecting cancer, particularly of the colon or
 PT prostate, comprises 3351 human polynucleotide sequences -

XX Claim 9: Page 722; 1046pp; English.

CC The present sequence is one of 3351 sequences in a library of human
 CC polynucleotides. The library is used to detect differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell and can
 CC detect colon, prostate, breast and lung cancer. The library can be used

CC to produce probes for detection of mRNA and to produce additional copies
 CC of the polynucleotides. The probes can be used for chromosome mapping of
 CC the polynucleotide and for detection of transcription levels. Ribozymes
 CC or antisense oligonucleotides can be generated. The polynucleotides and
 CC their gene products are used as genetic or biochemical markers (e.g. in
 CC blood or tissues) that will detect the earliest changes along the
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and
 CC preventive interventions. The polynucleotides, polypeptides and
 CC antibodies against them can be used in pharmaceutical compositions to
 CC treat the cancers and proliferative disorders such as neoplasia,
 CC dysplasia and hyperplasia.

SQ Sequence 336 BP; 97 A; 84 C; 59 G; 96 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 336;
 Best Local Similarity 100.0%; Pred. No. 80;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 atctgctgttgaa 22
 |||||
 Db 159 ATCTCTGTGTGAA 146

RESULT 15
 AAV87319/C
 ID AAV87319 standard; cDNA; 379 BP.

XX AAV87319;

XX 27-APR-1999 (first entry)

DE EST clone BP797.

XX Expressed sequence tag; secreted protein; haematopoiesis regulator;

KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;

KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;

XX receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

XX Homo sapiens.

XX WO9845435-A2.

XX 15-OCT-1998.

XX 10-APR-1998; 98WO-US06954.

XX 10-APR-1997; 97US-0835913.

PA (GENY) GENETICS INST INC.

PI Agostino MJ, Jacobs K, Lavallic FR, McCoy JM, Werberg D;

PI Racine LA, Spaulding V, Treacy M;

XX WPI: 1999-070076/06.

XX Claim 1: Page 527; 633pp; English.

CC This sequence represents an expressed sequence tag (EST), and is a
 CC polynucleotide of the invention. The polynucleotides of the invention are
 CC all secreted EST sequences isolated from a variety of human tissue
 CC sources. The EST sequences and proteins encoded by them are predicted to
 CC have useful biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals, although no supporting data is given. Suggested activities
 CC include nutritional activity, immune stimulating or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory

CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.

XX
SQ Sequence 379 BP; 124 A; 68 C; 70 G; 116 T; 1 other;

Query Match 56.0%; Score 14; DB 20; Length 379;

Best Local Similarity 100.0%; Pred. No. 81;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 gacatgctgttga 21
|||||

DB 129 GATCTGCTGTTTGA 116

RESULT 16

AAC26529/c ID AAC26529 standard; cDNA; 383 BP.

XX AAC26529;

AC 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 30604.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI; 2000-500381/45.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX
PS Claim 1; SEQ ID 30604; 71pp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.

XX
XX
SQ Sequence 383 BP; 96 A; 103 C; 63 G; 119 T; 2 other;

Query Match 56.0%; Score 14; DB 21; Length 383;

Best Local Similarity 100.0%; Pred. No. 81;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgattcgc 14
|||||

DB 263 GTTCTGATCTTC 250

RESULT 17

AAV01888/c ID AAV01888 standard; cDNA; 384 BP.

XX AAV01888;

DT 20-APR-1998 (first entry)

DE Human OTK27 gene.

KW Human; foetal brain cDNA library; GDP dissociation stimulating protein;

KW brain specific nucleosome assembly protein; diagnosis; therapy;

KW skeletal muscle specific ubiquitin conjugating enzyme; TMP-2; NPK;

KW nel-related protein type 1; nel-related type 2; hereditary disease;

XX cancer; OTK27; ss.

XX Homo sapiens.

OS

FT

FT

FT

PN EP796913-A2.

PD 24-SEP-1997.

PF 19-MAR-1997; 97EP-0104842.

XX 05-MAR-1997; 97JP-0069163.

PR 19-MAR-1996; 96JP-0063410.

XX (SAKA) OTSUKA PHARM CO LTD.

XX Fujiwara T, Horie M, Watanabe T;

PI WPI; 1997-459830/43.

DR P-PSDB; AAM37503.

XX Novel human genes, e.g. brain-specific nucleosome assembly protein -

PT useful for diagnosis or therapy of hereditary disease and cancer

XX Example 3; Page 35; 123pp; English.

PS The present sequence encodes a OTK27 isolated from a human foetal brain

CC cDNA library. The nucleotide or amino acid sequences are useful for

CC in-vitro diagnosis of hereditary diseases and cancer and for preparation

XX of pharmaceuticals.

XX
SQ Sequence 384 BP; 94 A; 111 C; 98 G; 81 T; 0 other;

Query Match 56.0%; Score 14; DB 18; Length 384;

Best Local Similarity 100.0%; Pred. No. 81;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tgatctgctgtt 19
|||||

DB 350 TCGATCTGCTGTTT 337

RESULT 18

AAC99862 ID AAC99862 standard; cDNA; 612 BP.

XX AAC99862;

DT 13-MAR-2001 (first entry)
 XX
 DE Human secreted protein gene 45 SEQ ID NO:55.
 XX
 KM Human: secreted protein; diagnosis; immunosuppressive; antiarthritic;
 KM antirheumatic; antiproliferative; cytostatic; cardiac; vasotropic;
 KM cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
 KM fungicide; ophthalmological; gene therapy; pathological condition;
 KM autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
 KM neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;
 KM cerebrovascular disorder; angiogenesis; nervous system disorder;
 KM Alzheimer's disease; infection; ocular disorder; corneal infection;
 KM wound healing; skin aging; food additive; preservative; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200070042-A1.
 XX
 PD 23-NOV-2000.
 XX
 PF 11-MAY-2000; 2000WO-US12788.
 XX
 PR 13-MAY-1999; 99US-0134068.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE;
 PI Duan RD, Florence KA, Soppet DR;
 XX
 DR WPI: 2000-679828/66.
 DR P-PSDB; AAB56121.
 XX
 PT Isolated nucleic acid molecule encoding a human secreted protein is
 PT used in preventing, treating or ameliorating a medical condition -
 XX
 XX Claim 1; Page 881; 1065pp; English.
 XX
 PS The polynucleotide sequences given in AAC99818 to AAC99977 encode the
 CC human secreted proteins given in AAB56077 to AAB56562. Human secreted
 CC proteins have activities based on the tissues and cells the genes are
 CC expressed in. Examples of activities include: immunosuppressive;
 CC antirheumatic; antiproliferative; antiproliferative; cardiac;
 CC vasotropic; cerebroprotective; neurotropic; cytostatic; antibacterial;
 CC virucide; fungicide; and ophthalmological. The human secreted
 CC polynucleotides and proteins can be used to prevent, treat or ameliorate
 CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
 CC dogs, chickens or sheep. They are also used in diagnosing a pathological
 CC condition or susceptibility to a pathological condition. Disorders which
 CC are diagnosed or treated include autoimmune diseases e.g. rheumatoid
 CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
 CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
 CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
 CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and
 CC fungi and ocular disorders e.g. corneal infection. The proteins can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The proteins can also be used as a
 CC food additive or preservative to increase or decrease storage
 CC capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used
 CC in the exemplification of the present invention.
 XX
 XX Sequence 612 BP; 199 A; 139 C; 139 G; 135 T; 0 other;
 XX
 Query Match 56.0%; Score 14; DB 21; Length 612;
 Best Local Similarity 100.0%; Pred. NO. 81;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 ttgctgacatgct 15
 Db 372 ttgctgacatgct 385

RESULT 19
 ID AAA44124/C
 XX AAA44124 standard; cDNA; 619 BP.
 XX
 AC AAA44124;
 XX
 DT 21-AUG-2000 (first entry)
 XX
 DE Human secreted expressed sequence tag SEQ ID NO:699.
 XX
 KM Human: mouse; chicken; rat; secreted expressed sequence tag; SEST;
 KM expressed sequence tag; EST; probe; chemotactic; proliferative;
 KM immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
 KM thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
 KM antiviral; antidiabetic; antiaslomatic; vulnery; antiparkinsonian;
 KM antidiabetic; osteopathic; neuroprotective; neurotropic; antiproliferative;
 KM cerebroprotective; anticonvulsant; antidepressant; gene therapy;
 KM vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
 KM insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
 KM lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
 KM central nervous system disorder; Alzheimer's disease; stroke;
 KM Parkinson's disease; Huntington's disease; coagulation disorder;
 KM haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
 KM tumour; infection; depression; psoriasis; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200021991-A1.
 XX
 PD 20-APR-2000.
 XX
 PF 15-OCT-1999; 99WO-US24206.
 XX
 PR 15-OCT-1998; 98US-0104436.
 XX
 PA (GEMY) GENETICS INST INC.
 XX
 PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M, Bowman MR;
 XX
 DR WPI: 2000-317938/27.
 XX
 PS Isolated polynucleotides, and encoded proteins, comprising secreted
 PT expressed sequence tags (SESTs), useful for treating various disorders
 PT such as autoimmune, infectious, and central nervous system disorders -
 XX
 XX Claim 1; Page 383; 803pp; English.
 XX
 PS AAA43426 to AAA45925 represent specifically claimed secreted expressed
 CC sequence tags (SESTs), isolated from human, mouse, chicken and rat
 CC tissue sources. The SESTs can have a range of activities depending on
 CC the tissues they were isolated from. The activities include:
 CC chemotactic; proliferative; immunomodulatory; haematopoietic;
 CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
 CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
 CC antiaslomatic; vulnery; antidiabetic; osteopathic; neuroprotective;
 CC neurotropic; antiparkinsonian; antiproliferative; cerebroprotective;
 CC anticonvulsant; and antidepressant. The SESTs can be used for gene
 CC therapy and in vaccines. The SESTs are useful as probes for the
 CC identification and isolation of full-length cDNAs and genomic DNA
 CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
 CC are useful in assays for determining biological activity and raising
 CC antibodies. They may be useful for treatment of autoimmune disorders
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
 CC osteoporosis, osteoarthritis, central nervous system disorders
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
 CC disease), tumours, bacterial, fungal or viral infections, depression and
 CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
 CC in the exemplification of the present invention.
 XX

SQ Sequence 619 BP; 193 A; 123 C; 125 G; 178 T; 0 other;

Query Match

Best Local Similarity 56.0%; Score 14; DB 21; Length 619;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 atcgcctgttga 22
| | | | | | | | | |
Db 43 ATCGCTGTTGAA 30

RESULT 20

AAFI4018/c
ID AAF14018 standard; cDNA: 707 BP.

XX AAF14018;

XX 13-MAR-2001 (first entry)

XX Aspergillus oryzae EST SEQ ID NO:6541.

XX Multiple gene expression: filamentous fungal cell; EST;
XX expressed sequence tag; Fusarium venenatum; Aspergillus niger;
XX Aspergillus oryzae; Trichoderma reesei; identification; recombination;
XX culture condition; environmental stress; spore morphogenesis;
XX metabolic pathway engineering; catabolic pathway engineering; ss.

XX Aspergillus oryzae.

XX MO200056762-A2.

XX 28-SEP-2000.

XX 22-MAR-2000; 2000MO-US07781.

XX 22-MAR-1999; 99US-0273623.

XX (NOVO) NOVO NORDISK BIOTECH INC.

XX (NOVO) NOVO NORDISK AS.

XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;

XX WPI; 2000-594572/56.

XX Monitoring differential expression of genes in filamentous fungal cells
XX uses fluorescence-labeled nucleic acids isolated from the cells and a
XX substrate of expressed sequence tags -

XX Claim 88; Page 2678; 3161pp; English.

XX The present invention describes a method for monitoring differential
XX expression of genes in a first filamentous fungal (FF) cell relative to
XX expression of the same genes in one or more second filamentous fungal
XX cells. The method uses fluorescence-labeled nucleic acids isolated from
XX the FF cells and a substrate of expressed sequence tags (EST). The ESTs
XX are used in the methods for monitoring differential expression of genes
XX in a first filamentous fungal (FF) cell relative to expression of the
XX same genes in one or more second filamentous fungal cells. Monitoring
XX the global expression of genes from FF cells allows the production
XX potential of the microorganisms to be improved. New genes may be
XX discovered, possible functions of unknown open reading frames can be
XX identified and gene copy number variation and stability can be
XX monitored. The expression of genes can be used to study how FF cells
XX adapt to changes in culture conditions, environmental stress, spore
XX morphogenesis, recombination, metabolic or catabolic pathway
XX engineering. Using ESTs provides several advantages over genomic or
XX random cDNA clones including elimination of redundancy as one spot on an
XX array equals one gene or open reading frame, and organisation of the
XX microarrays based on function of the gene products to facilitate
XX analysis of the results. AAF07478 to AAF11247 represents ESTs from
XX Fusarium venenatum; AAF11248 to AAF1853 represents ESTs from Aspergillus
XX niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and

CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.

SQ Sequence 707 BP; 154 A; 241 C; 156 G; 156 T; 0 other;

Query Match

Best Local Similarity 56.0%; Score 14; DB 21; Length 707;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 ctgagctgcgtgt 18
| | | | | | | | | |
Db 693 CTGAGCTGCTGTT 680

RESULT 21

AAQ92502
ID AAQ92502 standard; cDNA: 735 BP.

XX AAQ92502;

XX 07-FEB-1996 (first entry)

XX Mouse antibody F4-7 heavy chain variable region coding sequence.

XX Primer: amplification; PCR; mouse; kappa chain; heavy chain; Fab;
XX antibody; immunotolerance; animal; variegated display library;
XX variable region; antigen; immunorecessive; cell surface marker; foetal;
XX cancer; stem cell; variant; therapy; Alzheimer's disease; hybridoma;
XX familial hypercholesterolaemia; binding affinity; ds.

XX Mus musculus.

XX OS

XX FH

XX FT

XX FT

XX FT

XX PN

XX MO9515982-A2.

XX 15-JUN-1995.

XX 08-DEC-1994; 94MO-US14106.

XX 06-DEC-1994; 94US-0350400.

XX 08-DEC-1993; 93US-0164022.

XX (GENZ) GENZYME CORP.

XX Barsomian G, Copeland DP, Hillhouse D, Johnson T;

XX WPI; 1995-224291/29.

XX P-PSDB; AAR75458.

XX Generating new antibodies specific for immunorecessive epitopes -

XX by selection from variegated Y gene library cloned from

XX immuno-tolerance derived antibody repertoire, useful in diagnosis,

XX purification, and therapy, e.g. of cancer

XX Disclosure; Page 77-78; 1099p; English.

XX The coding sequence of the heavy chain variable region from the mouse
XX antibody F4-7. This sequence was isolated from a variegated display
XX library (VDL) of variable regions derived from a repertoire of
XX antibodies from an immunotolerised animal. The VDL is generated by PCR
XX amplifying the variable regions from the antibody coding sequences using
XX the primers AAQ74153-74. The variable regions, esp the complementarity
XX determining regions (CDR; see AAR75462-93 for examples of CDRs) from the
XX immunotolerant animals' antibodies are used to construct an antibody
XX against an immunorecessive antigen e.g. a cell surface marker on a foetal,
XX cancer or stem cell, which can differentiate between variant or related
XX forms of the antigen. The antibodies generated can be used in the
XX diagnosis, e.g. detection of the immunorecessive antigen, or in therapy

CC e.g. of cancer, Alzheimer's disease or familial hypercholesterolaemia.
 CC The method of production of the antibody allows rapid and sensitive
 CC isolation of antibodies that would be difficult to isolate by standard
 CC methods. The antibodies produced have greater binding affinity than
 CC those produced by combinatorial/hybridoma methods.

XX
 CC Sequence 735 BP; 172 A; 213 C; 188 G; 162 T; 0 other;

Query Match 56.0%; Score 14; DB 16; Length 735;
 Best Local Similarity 100.0%; Pred. No. 82;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 tctgtgattctgctg 16
 |||||
 Db 465 tctgtgattctgctg 478

RESULT 22

AA200347/C
 ID AA200347 standard; DNA; 753 BP.

XX
 AC AA200347;

XX
 DT 26-OCT-1999 (first entry)

XX
 DE Nucleotide sequence of human hsfATP3.

XX
 KW Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;
 KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.

XX
 OS Homo sapiens.

XX
 PN W09936537-A2.

XX
 PD 22-JUL-1999.

XX
 PF 14-JAN-1999; 99WO-US00182.

XX
 PR 14-JAN-1999; 99US-0232201.

XX
 PR 15-JAN-1998; 98US-0071374.

XX
 PR 20-JUL-1998; 98US-0093491.

XX
 PR 04-DEC-1998; 98US-0110941.

XX
 PR 14-JAN-1999; 99US-0232195.

XX
 PR 14-JAN-1999; 99US-0232197.

XX
 PR 14-JAN-1999; 99US-0232200.

XX
 PA (MILL-) MILLENNIUM PHARM INC.

XX
 PA (WHED) WHITEHEAD INST BIOMEDICAL RES.

XX
 PI Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;

XX
 DR WPI; 1999-444398/37.

XX
 DR P-PSDB; AAT14937.

XX
 PS Example 1; Fig 16; 255pp; English.

Fatty acid transport proteins and related polynucleotides, useful
 for treating obesity, diabetes and heart disease

CC The invention provides a family of fatty acid transport proteins (FATPs)
 CC that mediate transport of long chain fatty acids (LCFAs) across cell
 CC membranes into cells. Human and murine FATP proteins and nucleic acids
 CC encoding the proteins are provided. The FATP proteins can be produced
 CC by standard recombinant methodology. Fatty acid uptake by cells can be
 CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
 CC In particular, antisense oligonucleotides can be used to modulate FATP
 CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
 CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
 CC muscle or liver by administration of a complex of the agent and a FATP6
 CC binding moiety. DNA encoding FATP proteins can be used as a reference
 CC used in detecting variant alleles or homologues. Altering the LCFA uptake
 CC by administering an inhibitor or enhancer of FATP transport function in

CC the small intestine can decrease or increase calories available as fats,
 CC and can decrease or increase circulating fatty acids. Blocking the
 CC function of FATP4 and also FATP2, is useful for treating obesity,
 CC diabetes and heart disease.

XX
 CC Sequence 753 BP; 175 A; 206 C; 206 G; 166 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 753;
 Best Local Similarity 100.0%; Pred. No. 82;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 tctgtgttgaag 23
 |||||
 Db 436 TCTGCTGTTGAAG 423

RESULT 23

AAF15012/C
 ID AAF15012 standard; cDNA; 769 BP.

XX
 AC AAF15012;

XX
 DT 13-MAR-2001 (first entry)

XX
 DE Trichoderma reesei EST SEQ ID NO:7535.

XX
 KW Multiple gene expression; filamentous fungal cell; EST;

XX
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;

XX
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;

XX
 KW culture condition; environmental stress; spore morphogenesis;

XX
 KW metabolic pathway engineering; catabolic pathway engineering; ss.

XX
 OS Trichoderma reesei.

XX
 PN W0200056762-A2.

XX
 PD 28-SEP-2000.

XX
 PF 22-MAR-2000; 2000WO-US07781.

XX
 PR 22-MAR-1999; 99US-0273623.

XX
 PR (NOVO) NOVO NORDISK BIOTECH INC.

XX
 PR (NOVO) NOVO NORDISK AS.

XX
 PI Berka RM, Rey MW, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;

XX
 DR WPI; 2000-594572/56.

XX
 DR Monitoring differential expression of genes in filamentous fungal cells

XX
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a

XX
 PT substrate of expressed sequence tags -

XX
 PS Claim 89; Page 3043-3044; 3161pp; English.

CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an

CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AF07478 to AF11247 represents ESTs from
CC Fusarium venenatum; AF11248 to AF11853 represents ESTs from Aspergillus
CC niger; AF11854 to AF14878 represents ESTs from Aspergillus oryzae; and
CC AF14879 to AF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.

SO Sequence 769 BP; 203 A; 234 C; 203 G; 114 T; 15 other;

Query Match 56.0%; Score 14; DB 21; Length 769;

Best Local Similarity 100.0%; Pred. No. 82;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgcctgactcgtcg 16
|||||

Db 284 TGCTGATCTGCTG 271

RESULT 24

AAZ06233 standard; DNA; 891 BP.

AC AAZ06233;

DT 30-SEP-1999 (first entry)

DE Human secreted protein gene No. 15.

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
KM diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukemia;
KM developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KM immune system; aschma; lymphocytic disease; brain; hepatitis; lymphoma;
KM inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KM cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KM osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KM endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX Homo sapiens.

OS Homo sapiens.

PN WO935158-A1.

PD 15-JUL-1999.

XX 06-JAN-1999; 99WO-US00108.

XX 07-JAN-1998; 98US-0070704.

PR 07-JAN-1998; 98US-0070657.

PR 07-JAN-1998; 98US-0070658.

PR 07-JAN-1998; 98US-0070692.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Brewer LA, Duan RD, Ebner R, Lafleur DM, Ni J;

PI Olsen HS, Rosen CA, Ruben SM, Soppet DR;

DR WPI: 1999-44190/37.

XX P-PSDB; AAY38400, AAY38453.

XX New isolated human genes and the secreted polypeptides they encode

XX Claim 1; Page 158; 227pp; English.

CC This sequence represents a nucleic acid molecule which encodes a
CC secreted human protein. The gene number is given in the descriptor line.
CC The gene can be used to generate fusion proteins by linking to the gene
CC to a human immunoglobulin Fc portion (e.g. AAZ06210) for increasing the
CC stability of the fused protein as compared to the human protein only.
CC The invention relates to 36 novel genes and their fragments (nucleic
CC acid sequences: AAZ06219-206263; amino acid sequences AAY38386-38498)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological

CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 36
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAZ06219 for described uses).

SO Sequence 891 BP; 207 A; 214 C; 220 G; 250 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 891;

Best Local Similarity 100.0%; Pred. No. 82;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ttgcctgactcgtc 15
|||||

Db 727 ttgcctgactcgtc 740

RESULT 25

AAZ06234/C standard; DNA; 1044 BP.

AC AAZ06234;

DT 30-JUN-2000 (first entry)

DE L-asparaginase DNA.

XX L-asparaginase; fermentation; ss.

XX Unidentified.

OS CN1237633-A.

PN 08-DEC-1999.

XX 01-JUN-1998; 98CN-0102046.

XX 01-JUN-1998; 98CN-0102046.

XX (MICR-) INST MICROBIOLOGY CHINESE ACAD SCI.

XX Qian S, Wang Y, Meng G;

DR WPI: 2000-351194/31.

DR P-PSDB; AAM90867.

XX New structured L-asparaginase bacterial host cell and its production

XX culture -

XX Disclosure: Fig 2; 12pp; Chinese.

CC This invention describes a novel L-asparaginase-producing recombinant
CC bacteria and its fermentation culture method. The new cell is produced
CC using standard recombinant techniques. The enzyme gene expressed in
CC the cell is different from the reported enzyme gene in sequence, and
CC its individual amino acid composition is also different. The culture
CC medium adopted for culturing the new bacterial cells is cheap and easily
CC available, and the fermentation process used to grow the cells and
CC express the enzyme is simple and easy to implement. This sequence
CC encodes an L-asparaginase which is described in the method of the
CC invention.

SO Sequence 1044 BP; 274 A; 281 C; 272 G; 217 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1044;

Best Local Similarity 100.0%; Pred. No. 82;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgcctgactcgtcg 16
|||||

Db 1028 TGCTGATCTGCTG 1015

```

RESULT 26
AA209784
ID AA209784 standard; DNA; 1155 BP.
XX
AC AA209784;
XX
DT 23-NOV-1999 (first entry)
XX
DE E. coli S-adenosyl methionine synthase SAM DNA.
XX
KW SAM; S-adenosyl methionine synthase; biosynthesis; biotin; bios1; bios2;
KW bios3; cofactor; decarboxylation; Vitamin H; ss.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 1..1155
FT /tag= a
FT /product= "S-adenosyl methionine synthase"
FT
XX
PN DE19806872-A1.
XX
PD 26-AUG-1999.
XX
PF 19-FEB-1998; 98DE-1006872.
XX
PR 19-FEB-1998; 98DE-1006872.
XX
PA (BADI ) BASF AG.
XX
PI Schroeder H;
XX
DR WPI: 1999-480095/41.
DR P-PSDB; AAT33263.
XX
PT Production of biotin by expressing S-adenosyl-methionine synthase and
PT second biotin synthesis gene in host cells -
XX
PS Claim 1; Page 8-10; 48pp; German.
XX
CC This invention describes a novel method for the preparation of biotin
CC (1) which comprises expressing, in a prokaryotic or eukaryotic host
CC capable of producing (1): (a) an S-adenosyl-methionine synthase (SAM)
CC sequence (1), and (b) at least one of the other biotin biosynthesis
CC genes bios1, 2 or 3. (1) is a cofactor for enzyme-catalyzed
CC (de)carboxylation reactions and is an essential vitamin (Vitamin H) for
CC most animals and some microorganisms. Expression of biotin plus bios1,
CC bios2 or bios3 significantly increases productivity of biotin
CC biosynthesis, particularly by at least 3 times. This sequence encodes
CC the Escherichia coli SAM protein which is used in the method of the
CC invention.
XX
SQ Sequence 1155 BP; 277 A; 312 C; 303 G; 263 T; 0 other;

```

```

Query Match 56.0%; Score 14; DB 20; Length 1155;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 3 tgcgtgactgctg 16
   |||||||
DB 1031 tgcgtgactgctg 1044

```

```

RESULT 27
AAA99433
ID AAA99433 standard; DNA; 1155 BP.
XX
AC AAA99433;
XX
DT 26-JAN-2001 (first entry)

```

```

XX Met K oligonucleotide SEQ ID 17 used in preparation of L-methionine.
XX
XX L-methionine production; Met; Escherichia coli; ds.
XX
XX Escherichia coli.
XX
XX JP2000139471-A.
XX
XX 23-MAY-2000.
XX
XX 17-NOV-1998; 98JP-0326717.
XX
XX 17-NOV-1998; 98JP-0326717.
XX
XX (AJIN ) AJINOMOTO KK.
XX
XX WPI: 2001-018703/03.
XX
XX P-PSDB; AAB26814.
XX
XX Fermentative process for preparation of L-methionine (Met), comprises
XX using a modified Met producing microorganism, particularly an
XX Escherichia species
XX
XX Example 2; Page 17-18; 23pp; Japanese.
XX
XX The invention relates to a process for preparing L-methionine (Met)
XX using a modified Met producing organism. The modified microorganism has
XX its Met biosynthetic system repressor deleted, and has enhanced
XX homoserine transsuccinylase activity. The organism optionally has an
XX attenuated intracellular S-adenosyl methionine synthetase (SAM). The
XX process is used for the preparation of L-methionine. Sequences AAA99416
XX and AAA99433 encode proteins represented in AAB26816 and AAB26814
XX respectively, which are involved in the Escherichia coli Met biosynthetic
XX pathway. AAA99417-89941 (excluding AAA99433) represent primers, all
XX sequences are used in the production of the modified Met producing
XX organism of the invention.
XX
XX Sequence 1155 BP; 277 A; 312 C; 303 G; 263 T; 0 other;

```

```

Query Match 56.0%; Score 14; DB 22; Length 1155;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 3 tgcgtgactgctg 16
   |||||||
DB 1031 tgcgtgactgctg 1044

```

```

RESULT 28
AA200358/C
ID AA200358 standard; DNA; 1173 BP.
XX
AC AA200358;
XX
DT 26-OCT-1999 (first entry)
XX
DE Nucleotide sequence of human hSFATP3.
XX
KW Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;
KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
XX
OS Homo sapiens.
XX
XX WO9936537-A2.
XX
XX 22-JUL-1999.
XX
XX 14-JAN-1999; 99WO-US00182.
XX
XX 14-JAN-1999; 99US-0232201.
XX
XX 15-JAN-1998; 98US-0071374.

```


PR 20-JUL-1998; 98US-0093491.
PR 04-DEC-1998; 98US-0110941.
PR 14-JAN-1999; 99US-0232195.
PR 14-JAN-1999; 99US-0232197.
PR 14-JAN-1999; 99US-0232200.
XX
PA (MILL-) MILLENNIUM PHARM INC.
PA (WHEED) WHITEHEAD INST BIOMEDICAL RES.
XX
PI Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;
XX WPI: 1999-444398/37.
DR P-PSDB; AAY14948.
XX
PT Fatty acid transport proteins and related polynucleotides, useful
for treating obesity, diabetes and heart disease
XX
PS Claim 31; Fig 48; 255pp; English.
XX
CC The invention provides a family of fatty acid transport proteins (FATPs)
CC that mediate transport of long chain fatty acids (LCFAs) across cell
CC membranes into cells. Human and murine FATP proteins and nucleic acids
CC encoding the proteins are provided. The FATP proteins can be produced
CC by standard recombinant methodology. Fatty acid uptake by cells can be
CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
CC In particular, antisense oligonucleotides can be used to modulate FATP
CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
CC muscle or liver by administration of a complex of the agent and a FATP6
CC binding moiety. DNA encoding FATP proteins can be used as a reference
CC used in detecting variant alleles or homologues. Altering the LCFA uptake
CC by administering an inhibitor or enhancer of FATP transport function in
CC the small intestine can decrease or increase calories available as fats,
CC and can decrease or increase circulating fatty acids. Blocking the
CC function of FATP4 and also FATP2, is useful for treating obesity,
CC diabetes and heart disease.
XX
SQ Sequence 1173 BP; 247 A; 327 C; 351 G; 248 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 1173;
Best Local Similarity 100.0%; Pred. NO. 83;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tctgctgttggaag 23
|||||
DB 853 TCTGCTGTGGAAG 840

RESULT 29
AAC44861
ID AAC44861 standard; DNA; 1220 BP.
AC AAC44861;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 44409.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135352.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.

```

PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149425.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

```

```

PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158332.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      56.0%; Score 14; DB 21; Length 1220;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      3      tgcgtgacgcgcgcg 16
        |||
        317      tgcgtgacgcgcgcg 330

DB

RESULT      30
AAA81673/c
ID      AAA81673 standard; DNA; 1367 BP.
XX
AC      AAA81673;
XX
XX      04-DEC-2000 (first entry)
DT
XX
DE      N. meningitidis partial DNA sequence gnm_220 seq ID NO:220.
XX
KW      Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW      antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW      Meningococcus B; MenB; ds.
XX
XX      Neisseria meningitidis.
OS
PN      WO200022430-A2.
XX
PD      20-APR-2000.
XX
XX      08-OCT-1999; 99MO-US23373.
PF
XX
PR      09-OCT-1998; 98US-0103794.
PR      30-APR-1999; 99US-0132068.
XX
PA      (CHIR ) CHIRON CORP.
XX
XX      Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI      Masiangni V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI      Rappoli R, Pizza M;
XX

```

DR WPI: 2000-318079/27.
 XX Isolated nucleotide sequences of *Neisseria meningitidis* which can be
 PT used in the diagnosis and treatment of *N. meningitidis* infection and
 PT other *Neisseria* infections, for example, *N. gonorrhoea*.
 XX
 PS Claim 7: Page 1535; 1760pp; English.
 XX
 CC The present invention describes methods of obtaining immunogenic
 CC proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed *Neisseria meningitidis* genomic DNA
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC *Neisseria* DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to
 CC AAA81452 represent *Neisseria meningitidis* MemB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC *Neisseria* bacteria. For example, some of the identified proteins could
 CC be components of vaccines against *Meningococcus B*; against all serotypes;
 CC and/or against all pathogenic *Neisseria*. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC *Meningococcus B* vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.
 XX
 SQ Sequence 1367 BP; 412 A; 313 C; 290 G; 352 T; 0 other:
 Query Match 56.0%; Score 14; DB 21; Length 1367;
 Best Local Similarity 100.0%; Pred. No. 83;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 gctgagctgctgct 17
 |||||||||
 Db 1013 GCTGATCTGCTGT 1000
 RESULT 31
 AA01889/C
 ID AAV01889 standard; DNA; 1493 BP.
 XX
 AC AAV01889;
 XX
 DT 20-APR-1998 (first entry)
 XX
 DE Human OTK27 gene.
 XX
 DE Human OTK27 gene.
 XX
 KW Human: foetal brain cDNA library; GDP dissociation stimulating protein;
 KW brain specific nucleosome assembly protein; diagnosis; therapy;
 KW skeletal muscle specific ubiquitin conjugating enzyme; TMP-2; NP1K;
 KW nel-related protein type 1; nel-related type 2; hereditary disease;
 KW cancer; OTK27; ss.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH CDS 95..481
 FT /*tag= a
 FT /product= "OTK27"
 XX
 XX EP796913-A2.
 XX
 XX 24-SEP-1997.
 XX

PF 19-MAR-1997; 97EP-0104842.
 XX
 PR 05-MAR-1997; 97JP-0069163.
 PR 19-MAR-1996; 96JP-0063410.
 XX
 PA (SAKA) OTSUKA PHARM CO LTD.
 XX
 PI Fujiwara T, Horie M, Watanabe T;
 XX
 DR WPI: 1997-459830/43.
 DR P-PSDB; AAM37503.
 XX
 PR Novel human genes, e.g. brain-specific nucleosome assembly protein -
 PT useful for diagnosis or therapy of hereditary disease and cancer
 XX
 PS Example 3: Page 36-37; 123pp; English.
 XX
 CC The present sequence encodes a OTK27 isolated from a human foetal brain
 CC cDNA library. The nucleotide or amino acid sequences are useful for
 CC in-vitro diagnosis of hereditary diseases and cancer and for preparation
 CC of pharmaceuticals.
 XX
 SQ Sequence 1493 BP; 339 A; 363 C; 394 G; 397 T; 0 other:
 Query Match 56.0%; Score 14; DB 18; Length 1493;
 Best Local Similarity 100.0%; Pred. No. 83;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 tggactgctgtt 19
 |||||||||
 Db 444 TGGATCTGCTGTT 431

RESULT 32
 AAD03029/C
 ID AAD03029 standard; DNA; 1676 BP.
 XX
 AC AAD03029;
 XX
 DT 31-MAY-2001 (first entry)
 XX
 DE Flax 2S storage protein DNA.
 XX
 DE Flax; seed-specific promoter; storage protein; seed oil; vaccine;
 KW protein expression; anticoagulant; cytokine; growth factor; pectinase;
 KW interleukin; alpha-1-antitrypsin; anti-obesity protein; haemoglobin;
 KW serum albumin; insulin; lactoferrin; myoglobin; pulmonary surfactant;
 KW alpha amylase; ds.
 XX
 OS Linum usitatissimum.
 XX
 XX Key Location/Qualifiers
 FH primer_bind complement(1..18)
 FT /*tag= a
 FT /bound_moiety= "5' PCR primer(1)"
 FT repeat_unit 25..36
 FT /*tag= b
 FT /label= AT_rich_repeat
 FT repeat_unit 97..108
 FT /*tag= c
 FT /label= AT_rich_repeat
 FT repeat_unit 167..180
 FT /*tag= d
 FT /label= AT_rich_repeat
 FT repeat_unit 241..247
 FT /*tag= e
 FT /label= RV_like_repeat
 FT misc_signal 275..279
 FT /*tag= f
 FT /label= G-box
 FT misc_feature 285..290
 FT /*tag= g
 FT

```

FT      /note= "Seed specific box-like motif"
FT      327..333
FT      /tag= h
FT      primer_bind
FT      381..398
FT      /tag= 1
FT      /bound_moiety= "3' PCR primer(1)"
FT      418..942
FT      /tag= j
FT      /product= "Flax 2S storage protein"
FT      complement (943..957)
FT      /tag= k
FT      /bound_moiety= "5' PCR primer(2)"
FT      1158..1676
FT      /tag= 1
FT      /bound_moiety= "3' PCR primer(2)"
FT      W0200116340-A1.
PN      08-MAR-2001.
PD      25-AUG-2000; 2000WO-CA00988.
PF      27-AUG-1999; 99US-0151044.
PR      27-OCT-1999; 99US-0161722.
PR      30-MAY-2000; 2000CA-2310304.
XX      (SEMB-) SEMBIOSYS GENETICS INC.
XX      (CSTR) COMMONWEALTH SCI & IND RES ORG.
PI      Chaudhary S, Van Rooijen G, Moloney MM, Singh S;
XX      WPI: 2001-226693/23.
XX      P-PSDB; AAT72901.
DR      Expressing non-native genes in flax seeds and seeds of other plant
XX      species for altering the seed oil and protein composition in the seeds,
PT      comprises using seed-specific promoters obtained from flax
XX      Claim 14; Fig 3; 68pp; English.
PS      The present sequence is a DNA encoding flax 2S storage protein. The
XX      CC present invention relates to a method for expressing non-native genes in
XX      CC flax seeds. The method comprises introducing a chimeric nucleic acid
XX      CC construct containing a seed-specific promoter obtained from flax and a
XX      CC nucleic acid which is non-native to the promoter, into a flax plant cell
XX      CC and growing the plant cell into a mature flax plant capable of setting
XX      CC seed, where the nucleic acid is expressed in the seed under the control
XX      CC of the promoter. The seed-specific promoters obtained from flax are
XX      CC useful for modifying the protein, oil or polysaccharide composition of
XX      CC the flax seeds and seeds of other plant species. The promoters facilitate
XX      CC expression of proteins, including sulphur-rich protein that are found in
XX      CC lupinus or Brazil nuts in a seed deficient in sulphurous amino acids,
XX      CC peptides having pharmaceutical value such as anticoagulants, antibodies,
XX      CC vaccines, cytokines, growth factors, interleukins, mammalian proteins,
XX      CC including alpha-1-antitrypsin, anti-obesity proteins, haemoglobin, blood
XX      CC proteins, human serum albumin, insulin, lactoferrin, myoglobin, pulmonary
XX      CC surfactants and proteins of industrial value such as alpha-amylase,
XX      CC arabinase, amyloglucosidase, catalase, cellobiohydrolase, pectinases,
XX      CC phytase, papain and xylanase.
XX      Sequence 1676 BP; 516 A; 367 C; 442 G; 351 T; 0 other;
SQ

```

```

Query Match          56.0%; Score 14; DB 22; Length 1676;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3 tgcctgactctgctg 16
        |||
Db      584 TGCTGATCTGCTG 571

```

RESULT 33

```

AAT96346/c
ID      AAT96346 standard; cDNA; 1848 BP.
XX
AC      AAT96346;
XX
DT      08-APR-1998 (first entry)
XX
DE      Chimeric gene containing anti-asparaginase Mab light and heavy chain.
XX
XX      Immunoglobulin; Ig; heavy chain; variable region; murine; human;
XX      asparaginase II; monoclonal antibody; Mab; light chain;
XX      recombinant chimeric polypeptide; ss.
XX
OS      Synthetic.
OS      Chimeric - Homo sapiens.
OS      Chimeric - Mus sp.
XX
PN      US5686579-A.
PD      11-NOV-1997.
XX
XX      23-MAY-1995; 95US-0447422.
XX
XX      22-JUN-1993; 93US-0081410.
XX      21-JUN-1988; 88US-0205748.
XX      31-AUG-1992; 92US-0938505.
XX      23-MAY-1995; 95US-0447422.
XX
XX      (HYBR-) HYBRISSENS LTD.
XX
PI      Ramjeesingh M, Rothstein A, Shami EY;
XX      WPI: 1997-558200/51.
XX
DR      Self-protecting chimeric polypeptide comprising biologically active
XX      PT sequence and single-chain antibody sequence - has resistance to e.g.
XX      PT disrupting temperature, presence of proteolytic enzymes, etc.
XX
XX      Example 2; Columns 27-30; 29pp; English.
PS      The present sequence is a chimeric gene containing the cDNA for
XX      CC the immunoglobulin (Ig) heavy and light chain variable regions of a
XX      CC murine anti-asparaginase II monoclonal antibody (Mab), and human
XX      CC asparaginase II. The gene was used in the preparation of a novel
XX      CC recombinant chimeric polypeptide, comprising a 1st region
XX      CC comprising a biologically active domain and another domain
XX      CC containing an epitope, linked via a polypeptide to a 2nd region
XX      CC including a single chain antibody (SCA) having the light and heavy
XX      CC chains of an antibody variable region which specifically binds the
XX      CC epitope in the 1st region. The chimeric polypeptide assumes a
XX      CC conformation in which the SCA is bound to the epitope of the 1st
XX      CC region and protects its biological activity from deactivation by
XX      CC denaturing temperatures or pH conditions, proteolytic enzymes,
XX      CC oxidising agents or alcohol. The regions of the chimeric
XX      CC polypeptide interact to form a structure analogous to an
XX      CC antibody-antigen complex. A L-asparaginase-SCA fusion protein of
XX      CC the above type has better trypsin resistance than free
XX      CC L-asparaginase.
XX      Sequence 1848 BP; 435 A; 498 C; 517 G; 398 T; 0 other;
SQ

```

```

Query Match          56.0%; Score 14; DB 18; Length 1848;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3 tgcctgactctgctg 16
        |||
Db      1829 TGCTGATCTGCTG 1816

```

RESULT 34
AAZ00367/c

```

ID  AA200367 standard; DNA; 1998 BP.
XX
AC  AA200367;
XX
DT  26-OCT-1999 (first entry)
XX
DE  Nucleotide sequence of murine mmFATP3.
XX
KW  Fatty acid transport protein; FATP; long chain fatty acid; LCFA; murine;
XX  fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
XX
OS  Mus sp.
XX
PN  W09936537-A2.
XX
PD  22-JUL-1999.
XX
PF  14-JAN-1999; 99WO-US00182.
XX
PR  14-JAN-1999; 99US-0232201.
XX  15-JAN-1998; 98US-0071374.
XX  20-JUL-1998; 98US-0093491.
XX  04-DEC-1998; 98US-0110941.
XX  14-JAN-1999; 99US-0232195.
XX  14-JAN-1999; 99US-0232197.
XX  14-JAN-1999; 99US-0232200.
XX
PA  (MILL-) MILLENNIUM PHARM INC.
PA  (WHEB) WHITEHEAD INST BIOMEDICAL RES.
XX
PI  Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;
XX
DR  WPI: 1999-44398/37.
XX  P-PSDB; AAY14957.
XX
PT  Fatty acid transport proteins and related polynucleotides, useful
XX  for treating obesity, diabetes and heart disease
XX
PS  Example 1; Fig 66; 255pp; English.
XX
CC  The invention provides a family of fatty acid transport proteins (FATPs)
CC  that mediate transport of long chain fatty acids (LCFAs) across cell
CC  membranes into cells. Human and murine FATP proteins and nucleic acids
CC  encoding the proteins are provided. The FATP proteins can be produced
CC  by standard recombinant methodology. Fatty acid uptake by cells can be
CC  modulated by modulating biosynthesis of FATP proteins especially FATP6.
CC  In particular, antisense oligonucleotides can be used to modulate FATP
CC  biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
CC  uptake in cardiac muscle of humans. Agents can be directed to cardiac
CC  muscle or liver by administration of a complex of the agent and a FATP6
CC  binding moiety. DNA encoding FATP proteins can be used as a reference
CC  used in detecting variant alleles or homologues. Altering the LCFA uptake
CC  by administering an inhibitor or enhancer of FATP transport function in
CC  the small intestine can decrease or increase calories available as fats,
CC  and can decrease or increase circulating fatty acids. Blocking the
CC  function of FATP4 and also FATP2, is useful for treating obesity,
CC  diabetes and heart disease.
XX
SQ  Sequence 1998 BP; 422 A; 567 C; 597 G; 412 T; 0 other;

Query Match          56.0%; Score 14; DB 20; Length 1998;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  10 tctgctgttgaag 23
    |||||
DB  1690 TCTGCTGTTGAAG 1677

RESULT 35
AAZ30162
ID  AAZ30162 standard; cDNA; 2080 BP.

```

```

XX
AC  AAZ30162;
XX
DT  26-JAN-2000 (first entry)
XX
DE  cDNA encoding a portion of the starch R1 phosphorylation protein.
XX
KW  Starch R1 phosphorylation protein; starch degradation;
XX  plant starch biosynthesis; ss.
XX
OS  Glycine max.
XX
FH  Key
XX  CDS
XX  Location/Qualifiers
XX  3..1748
XX  FT
XX  CDS
XX  /tag= a

PN  W09953072-A1.
XX
PD  21-OCT-1999.
XX
PF  08-APR-1999; 99WO-US07639.
XX
PR  09-APR-1998; 98US-0081143.
XX
PA  (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI  Cressman RF, Allen SM;
XX
DR  WPI: 1999-620435/53.
XX  P-PSDB; AAY43629.
XX
PT  New starch R1 phosphorylation protein homologues
XX
PS  Claim 2; Page 41-42; 54pp; English.
XX
CC  The present sequence encodes a starch R1 phosphorylation protein.
CC  The protein is involved in starch degradation. The specification
CC  also describes a chimeric gene encoding all or a portion of the
CC  starch R1 phosphorylation protein, in the sense or antisense
CC  orientation, where expression of the chimeric gene results in
CC  production of altered levels of starch R1 phosphorylation protein
CC  in transformed plants or cells. The protein facilitates studies of
CC  starch degradation in plants, and provides tools for the genetic
CC  manipulation of starch biosynthesis in plants.
XX
SQ  Sequence 2080 BP; 623 A; 385 C; 467 G; 605 T; 0 other;

Query Match          56.0%; Score 14; DB 20; Length 2080;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  3 tgcgtgactgctg 16
    |||||
DB  290 tgcgtgactgctg 303

RESULT 36
AAZ00343/C
ID  AAZ00343 standard; DNA; 2087 BP.
XX
AC  AAZ00343;
XX
DT  26-OCT-1999 (first entry)
XX
DE  Nucleotide sequence of murine mmFATP3.
XX
KW  Fatty acid transport protein; FATP; long chain fatty acid; LCFA; murine;
XX  fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
XX
OS  Mus musculus.
XX
PN  W09936537-A2.

```

```

XX 22-JUL-1999.
PD
XX
XX 14-JAN-1999. 99WO-US00182.
PF
XX 14-JAN-1999. 99US-0232201.
PR 15-JAN-1998. 98US-0071374.
PR 20-JUL-1998. 98US-0093491.
PR 04-DEC-1998. 98US-0110941.
PR 14-JAN-1999. 99US-0232195.
PR 14-JAN-1999. 99US-0232197.
PR 14-JAN-1999. 99US-0232200.
XX
XX (MILL-) MILLENNIUM PHARM INC.
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
PI Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;
XX
XX WPI: 1999-444398/37.
DR P-PSDB; AAY14933.
XX
XX Fatty acid transport proteins and related polynucleotides, useful
PT for treating obesity, diabetes and heart disease
XX
XX Example 1; Fig 8; 255pp; English.
PS
XX The invention provides a family of fatty acid transport proteins (FATPs)
CC that mediate transport of long chain fatty acids (LCFAs) across cell
CC membranes into cells. Human and murine FATP proteins and nucleic acids
CC encoding the proteins are provided. The FATP proteins can be produced
CC by standard recombinant methodology. Fatty acid uptake by cells can be
CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
CC In particular, antisense oligonucleotides can be used to modulate FATP
CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
CC muscle or liver by administration of a complex of the agent and a FATP6
CC binding moiety. DNA encoding FATP proteins can be used as a reference
CC used in detecting variant alleles or homologues. Altering the LCFA uptake
CC by administering an inhibitor or enhancer of FATP transport function in
CC the small intestine can decrease or increase calories available as fats,
CC and can decrease or increase circulating fatty acids. Blocking the
CC function of FATP4 and also FATP2, is useful for treating obesity,
CC diabetes and heart disease.
XX
XX Sequence 2087 BP; 441 A; 596 C; 624 G; 426 T; 0 other;
SQ
Query Match 56.0%; Score 14; DB 20; Length 2087;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 10 lctgctgtttgaag 23
DB 1779 TCTGCTGTTGAAG 1766
|||||

```

```

PD 22-JUL-1999.
XX
XX 14-JAN-1999. 99WO-US00182.
PF
XX 14-JAN-1999. 99US-0232201.
PR 15-JAN-1998. 98US-0071374.
PR 20-JUL-1998. 98US-0093491.
PR 04-DEC-1998. 98US-0110941.
PR 14-JAN-1999. 99US-0232195.
PR 14-JAN-1999. 99US-0232197.
PR 14-JAN-1999. 99US-0232200.
XX
XX (MILL-) MILLENNIUM PHARM INC.
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
PI Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;
XX
XX WPI: 1999-444398/37.
DR P-PSDB; AAY14969.
XX
XX Fatty acid transport proteins and related polynucleotides, useful
PT for treating obesity, diabetes and heart disease
XX
XX Claim 31; Fig 94A-B; 255pp; English.
PS
XX The invention provides a family of fatty acid transport proteins (FATPs)
CC that mediate transport of long chain fatty acids (LCFAs) across cell
CC membranes into cells. Human and murine FATP proteins and nucleic acids
CC encoding the proteins are provided. The FATP proteins can be produced
CC by standard recombinant methodology. Fatty acid uptake by cells can be
CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
CC In particular, antisense oligonucleotides can be used to modulate FATP
CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
CC muscle or liver by administration of a complex of the agent and a FATP6
CC binding moiety. DNA encoding FATP proteins can be used as a reference
CC used in detecting variant alleles or homologues. Altering the LCFA uptake
CC by administering an inhibitor or enhancer of FATP transport function in
CC the small intestine can decrease or increase calories available as fats,
CC and can decrease or increase circulating fatty acids. Blocking the
CC function of FATP4 and also FATP2, is useful for treating obesity,
CC diabetes and heart disease.
XX
XX Sequence 2166 BP; 370 A; 682 C; 695 G; 419 T; 0 other;
SQ
Query Match 56.0%; Score 14; DB 20; Length 2166;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 10 lctgctgtttgaag 23
DB 1987 TCTGCTGTTGAAG 1974
|||||

```

```

RESULT 37
AAZ00379/c
ID AAZ00379 standard; DNA; 2166 BP.
XX
XX AAZ00379;
AC
XX 26-OCT-1999 (first entry)
DT
XX Nucleotide sequence of human hsfATP5 gene.
DE
XX Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;
KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
XX
XX Homo sapiens.
OS
XX
XX W09936537-A2.
XX
XX

```

```

RESULT 38
AAF27417/c
ID AAF27417 standard; cDNA; 2405 BP.
XX
XX AAF27417;
AC
XX 24-APR-2001 (first entry)
DT
XX Human fatty acid transporter PSEC67 cDNA.
DE
XX Human; fatty acid transporter; PSEC67; long-chain fatty acid uptake;
KW oleic acid; drug screening; gene therapy; metabolic disorder;
KW cardiomyopathy; skeletal muscle disorders; renal failure; ss.
XX
XX Homo sapiens.
OS
XX
XX W0200104301-A1.
XX
XX

```

PD 18-JAN-2001.
XX
PF 07-JUL-2000: 2000MO-JP04549.
XX
PR 08-JUL-1999: 99JP-0194179.
PR 18-OCT-1999: 99US-0159586.
PR 25-APR-2000: 2000JP-0128993.
XX
PA (HELI-) HELIX RES INST.
XX
PI Morikawa N, Masuho Y, Ota T, Isogai T, Nishikawa T, Kawai Y;
XX
DR WPI: 2001-138349/14.
XX
DR P-PDB: AAB60388.
XX
PT Fatty acid transporter protein and encoded gene PSEC67 cloned from
PT human cDNA library, with activity of oleic acid incorporation, useful
PT as target molecule of preventives or remedies of fatty-acid metabolic
PT disorders -
XX
PS Claim 1: Page 44-48; 58pp; Japanese.
XX
CC The invention relates to a novel human fatty acid transporter,
CC PSEC67 (AAB60388), and to cDNA encoding it (AA27417). PSEC67 is
CC responsible for the uptake of oleic acid into cells. The invention
CC also relates to vectors and host cells comprising a PSEC67
CC nucleic acid: the recombinant production of PSEC67; an antibody
CC against PSEC67; methods of screening for compounds which can regulate
CC the uptake of long-chain fatty acids into cells; and the compounds thus
CC identified. The PSEC67 protein and the gene encoding it are useful as
CC targets for the treatment or prevention of diseases associated with
CC defective metabolism of long-chain fatty acids. Such diseases include
CC cardiomyopathy, skeletal muscle disorders, or renal failure. PSEC67
CC nucleic acids may also be used in gene therapy for such disorders.
CC The present sequence represents cDNA encoding PSEC67.
XX
SQ Sequence 2405 BP; 410 A; 735 C; 781 G; 479 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 2405;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tctgtctgttggaag 23
Db 2111 TCTGCTCTTGAAG 2098

RESULT 39
AA233977/C
ID AA233977 standard; cDNA; 2574 BP.
XX
AC AA233977;
XX
DT 07-DEC-1999 (first entry)
XX
DE Human PRO703 nucleotide sequence.
XX
KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW secreted protein; transmembrane protein; ss.
XX
OS Homo sapiens.
XX
PN WO946281-A2.
XX
PD 16-SEP-1999.
XX
PF 08-MAR-1999: 99WO-US05028.
XX
PR 10-MAR-1998: 98US-0077450.
PR 11-MAR-1998: 98US-0077632.
PR 11-MAR-1998: 98US-0077641.

PR 11-MAR-1998: 98US-0077649.
PR 12-MAR-1998: 98US-0077791.
PR 13-MAR-1998: 98US-0078004.
PR 17-MAR-1998: 98US-0040220.
PR 20-MAR-1998: 98US-0078886.
PR 20-MAR-1998: 98US-0078910.
PR 20-MAR-1998: 98US-0078936.
PR 20-MAR-1998: 98US-0078939.
PR 25-MAR-1998: 98US-0079294.
PR 26-MAR-1998: 98US-0079636.
PR 27-MAR-1998: 98US-0079663.
PR 27-MAR-1998: 98US-0079664.
PR 27-MAR-1998: 98US-0079689.
PR 27-MAR-1998: 98US-0079728.
PR 27-MAR-1998: 98US-0079786.
PR 30-MAR-1998: 98US-0079920.
PR 30-MAR-1998: 98US-0079923.
PR 31-MAR-1998: 98US-0080105.
PR 31-MAR-1998: 98US-0080107.
PR 31-MAR-1998: 98US-0080165.
PR 31-MAR-1998: 98US-0080194.
PR 01-APR-1998: 98US-0080327.
PR 01-APR-1998: 98US-0080328.
PR 01-APR-1998: 98US-0080333.
PR 01-APR-1998: 98US-0080334.
PR 08-APR-1998: 98US-0081049.
PR 08-APR-1998: 98US-0081070.
PR 08-APR-1998: 98US-0081071.
PR 09-APR-1998: 98US-0081195.
PR 09-APR-1998: 98US-0081203.
PR 09-APR-1998: 98US-0081229.
PR 15-APR-1998: 98US-0081817.
PR 15-APR-1998: 98US-0081838.
PR 15-APR-1998: 98US-0081952.
PR 15-APR-1998: 98US-0081955.
PR 21-APR-1998: 98US-0082568.
PR 21-APR-1998: 98US-0082569.
PR 22-APR-1998: 98US-0082700.
PR 22-APR-1998: 98US-0082704.
PR 22-APR-1998: 98US-0082804.
PR 23-APR-1998: 98US-0082767.
PR 23-APR-1998: 98US-0082796.
PR 27-APR-1998: 98US-0083336.
PR 28-APR-1998: 98US-0083322.
PR 29-APR-1998: 98US-0083392.
PR 29-APR-1998: 98US-0083495.
PR 29-APR-1998: 98US-0083496.
PR 29-APR-1998: 98US-0083499.
PR 29-APR-1998: 98US-0083500.
PR 29-APR-1998: 98US-0083545.
PR 29-APR-1998: 98US-0083554.
PR 29-APR-1998: 98US-0083558.
PR 30-APR-1998: 98US-0083742.
PR 05-MAY-1998: 98US-0083742.
PR 05-MAY-1998: 98US-0084366.
PR 06-MAY-1998: 98US-0084414.
PR 06-MAY-1998: 98US-0084441.
PR 07-MAY-1998: 98US-0084598.
PR 07-MAY-1998: 98US-0084600.
PR 07-MAY-1998: 98US-0084627.
PR 07-MAY-1998: 98US-0084637.
PR 07-MAY-1998: 98US-0084639.
PR 07-MAY-1998: 98US-0084640.
PR 07-MAY-1998: 98US-0084643.
PR 13-MAY-1998: 98US-0085323.
PR 13-MAY-1998: 98US-0085338.
PR 13-MAY-1998: 98US-0085339.
PR 15-MAY-1998: 98US-0085573.
PR 15-MAY-1998: 98US-0085579.
PR 15-MAY-1998: 98US-0085580.
PR 15-MAY-1998: 98US-0085582.
PR 15-MAY-1998: 98US-0085689.
PR 15-MAY-1998: 98US-0085697.

PR 15-MAY-1998: 98US-0085700.
 PR 15-MAY-1998: 98US-0085704.
 PR 18-MAY-1998: 98US-0086023.
 PR 22-MAY-1998: 98US-0086392.
 PR 22-MAY-1998: 98US-0086414.
 PR 22-MAY-1998: 98US-0086430.
 PR 22-MAY-1998: 98US-0086486.
 PR 28-MAY-1998: 98US-0087106.
 PR 28-MAY-1998: 98US-0087208.
 PR 30-JUL-1998: 98US-0094651.
 PR 11-SEP-1998: 98US-0100038.
 PR XX
 PA (GETH) GENENTECH INC.
 XX
 PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
 XX
 DR WPI, 1999-551358/46.
 DR P-PSDB; AAY41699.
 XX
 PT New secreted and transmembrane polypeptides and their polynucleotides,
 PT useful for treating blood coagulation disorders, cancers and cellular
 PT adhesion disorders -
 XX
 PS Claim 2; Fig 38; 530pp; English.
 XX
 CC The present invention describes secreted and transmembrane polypeptides
 CC and their polynucleotides. The nucleotide sequences are useful as
 CC sources of probes, primers, for chromosome mapping, and for generation
 CC of antisense sequences. They can also be used to create transgenic
 CC animals. The proteins can be used to treat a variety of diseases and
 CC disorders, depending on their function. Diseases that may be treated
 CC include blood coagulation disorders, cancers and cellular adhesion
 CC disorders. They may also be used to raise antibodies. AA233891 to
 CC AA234338, and AAY41685 to AAY41774 represent polynucleotide and
 CC polypeptide sequence given in the exemplification of the present
 CC invention.
 CC
 SQ Sequence 2574 BP; 470 A; 775 C; 821 G; 508 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 2574;
 Best Local Similarity 100.0%; Pred. No. 84;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tctgctgttgaag 23
 |||||||||
 Db 2167 TCTGCTGTTGAAG 2154

RESULT 40
 AAC78481/C
 ID AAC78481 standard; cDNA; 2574 BP.
 XX
 AC AAC78481;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human PRO703 (UNQ367) nucleotide sequence SEQ ID NO:101.
 XX
 KW Human: secreted protein; transmembrane protein; PRO; EST; cytosolic;
 KW expressed sequence tag; detection; cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200053756-A2.
 XX
 PD 14-SEP-2000.
 XX
 PE 18-FEB-2000; 2000WO-US04341.
 XX
 PF 08-MAR-1999; 99WO-US05028.
 PR 12-MAR-1999; 99US-0123957.

PR 29-MAR-1999; 99US-0126773.
 PR 21-APR-1999; 99US-0130232.
 PR 28-APR-1999; 99US-0131445.
 PR 14-MAY-1999; 99US-0134287.
 PR 23-JUN-1999; 99US-0141037.
 PR 26-JUL-1999; 99US-0145698.
 PR 29-OCT-1999; 99US-0162506.
 PR 30-NOV-1999; 99WO-US28313.
 PR 02-DEC-1999; 99WO-US28551.
 PR 02-DEC-1999; 99WO-US28565.
 PR 16-DEC-1999; 99WO-US30095.
 PR 30-DEC-1999; 99WO-US31243.
 PR 30-DEC-1999; 99WO-US31274.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 06-JAN-2000; 2000WO-US00277.
 PR 06-JAN-2000; 2000WO-US00376.
 PR XX
 PA (GETH) GENENTECH INC.
 XX
 PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL,
 PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gertsen ME,
 PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ,
 PI Klyavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA,
 PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
 XX
 DR WPI: 2000-611443/58.
 DR P-PSDB; AAB44255.
 XX
 PT Novel PRO polypeptides and polynucleotides used in detection methods,
 PT to target bioactive molecules to specific cells, and to modulate
 PT cellular activities -
 XX
 PS Claim 2; Fig 38; 636pp; English.
 XX
 CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed
 CC sequence tag) sequences which encode secreted or transmembrane PRO
 CC polypeptides. The PRO polynucleotides and polypeptides have cytostatic
 CC activity. The polynucleotides and polypeptides can be used for detecting
 CC the presence of PRO polypeptides in samples, for linking bioactive
 CC molecules to cells and for modulating biological activities of cells,
 CC using the polypeptides for specific targeting. The polypeptide targeting
 CC can be used to kill the target cells, e.g. for the treatment of cancers.
 CC The polypeptide pairs provide specific targeting of bioactive molecules
 CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
 CC the isolation of the PRO polynucleotide sequences.
 CC
 SQ Sequence 2574 BP; 470 A; 775 C; 821 G; 508 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 2574;
 Best Local Similarity 100.0%; Pred. No. 84;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tctgctgttgaag 23
 |||||||||
 Db 2167 TCTGCTGTTGAAG 2154

RESULT 41
 AAC58239/C
 ID AAC58239 standard; cDNA; 2574 BP.
 XX
 AC AAC58239;
 XX
 DT 25-JAN-2001 (first entry)
 XX
 DE Human PRO703 nucleotide sequence SEQ ID NO:28.
 XX
 KW Human: tumour; diagnosis; neoplastic disease; identification; cancer;
 KW tumorigenesis; detection; neoplastic cell growth; proliferation;
 KW cytostatic; antiinflammatory; immunomodulatory; inflammatory disorder;
 KW immunological disorder; ss.
 XX

OS Homo sapiens.
 XX
 PN W0200053754-A1.
 XX
 PD 14-SEP-2000.
 XX
 PF 06-JAN-2000; 2000MO-US00277.
 XX
 PR 08-MAR-1999; 99MO-US05028.
 PR 12-MAR-1999; 99US-0123957.
 PR 29-MAR-1999; 99US-0126773.
 PR 21-APR-1999; 99US-0130232.
 PR 28-APR-1999; 99US-0131445.
 PR 05-OCT-1999; 99MO-US23089.
 PR 30-NOV-1999; 99MO-US28313.
 PR 02-DEC-1999; 99MO-US28551.
 PR 02-DEC-1999; 99MO-US28564.
 PR 30-DEC-1999; 99MO-US31243.
 PR 30-DEC-1999; 99MO-US31274.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Baker KP, Desauvage FJ, Goddard A, Gurney AL, Klein RD, Roy MA;
 PI Wood WI;
 XX
 DR WPI: 2000-572269/53.
 DR P-PSDB: AAB24054.
 XX
 PT New isolated antibody for use in compositions and methods for the
 PT diagnosis and treatment of neoplastic cell growth and proliferation in
 PT mammals, including humans, and in monitoring tumor treatment -
 XX
 PS Claim 50; Fig 28; 195pp; English.
 XX
 CC The present invention describes an isolated antibody (Ab) that binds to
 CC one of the human proteins (P) designated PRO213, PRO1330, PRO1449,
 CC PRO327, PRO324, PRO351, PRO362, PRO615, PRO531, PRO364, PRO618,
 CC PRO772, PRO703, PRO792 or PRO474. The Ab can be used in compositions
 CC and methods for the diagnosis and treatment of neoplastic cell growth
 CC and proliferation in mammals, including humans. Genes and polypeptides
 CC encoded by them, that are amplified in the genome of a tumour cell, can
 CC be identified and are useful targets for the treatment and prevention of
 CC certain cancers and may be used to monitor tumor treatment. Compounds
 CC that inhibit the expression or activity of the identified polypeptides
 CC can be identified and used as antagonists. Benign or malignant tumours,
 CC inflammatory disorders and immunological disorders can be treated.
 CC AAC58123 to AAC58224 represent hybridisation probes and PCR primers used
 CC in the isolation of the human PRO sequences. AAC58225 to AAC58241 and
 CC AAB24041 to AAB24056 represent human PRO polynucleotide and protein
 CC sequences given in the exemplification of the present invention.
 XX
 SQ Sequence 2574 BP; 470 A; 775 C; 821 G; 508 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 2574;
 Best Local Similarity 100.0%; Pred. No. 84;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 ctgcgtgttgaagc 23
 |||
 DB 2167 TCTGCTGTTGAAG 2154

RESULT 42
 AA225020/C
 ID AA225020 standard; cDNA; 2669 BP.
 XX
 AC AA225020;
 XX
 DT 07-DEC-1999 (first entry)
 XX
 DE Human H-beta 58 family CBLAC05 nucleotide sequence.
 XX

KW Human: H-beta 58 family: CBLAC05; cancer; AIDS; metabolic disorder;
 KW neurological disease; developmental abnormality; diagnosis;
 KW spontaneous abortion; ss.
 XX
 OS Homo sapiens.
 XX
 PF Key Location/Qualifiers
 FT CDS 80..1063
 FT /*tag= a
 FT /product= "CBLAC05"
 FT /note= "an H-beta 58 family protein"
 XX
 PN W09947663-A1.
 XX
 PD 23-SEP-1999.
 XX
 PF 18-MAR-1998; 98MO-CN00042.
 XX
 PR 18-MAR-1998; 98MO-CN00042.
 XX
 PA (YUSH-) UNIV SHANGHAI SECOND MEDICAL.
 XX
 PI Zhang Q, Kan L, He K, Shen Y;
 XX
 DR WPI: 1999-571836/48.
 DR P-PSDB: AAY41679.
 XX
 PT CBLAC05, a homologue of murine H-beta 58, and related polynucleotides
 PT
 PS Claim 4; Page 9-11; 33pp; English.
 XX
 CC The present sequence encodes a human H-beta 58 family protein designated
 CC CBLAC05. CBLAC05 is homologous to the murine H-beta 58 protein. The
 CC CBLAC05 polynucleotide (1), vectors containing (1) and recombinant host
 CC cells are useful for recombinant production of CBLAC05. Both (1),
 CC CBLAC05 and antibodies against CBLAC05 are useful as research reagents,
 CC for screening assays and in diagnostic assays. Antagonists and agonists
 CC of CBLAC05 can be used to inhibit or enhance, respectively, the activity
 CC of CBLAC05 or expression of (1). Anti-CBLAC05 antibodies and CBLAC05
 CC or its fragments can be used in vaccines. In particular, the proteins,
 CC antibodies, agonists and antagonists can be used for treating, e.g.
 CC cancer, AIDS, metabolic disorders, neurological disease, developmental
 CC abnormalities and spontaneous abortions, related to both an excess and
 CC insufficient amounts of CBLAC05.
 XX
 SQ Sequence 2669 BP; 864 A; 419 C; 530 G; 856 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 2669;
 Best Local Similarity 100.0%; Pred. No. 84;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 ctgcgtgttgaagc 24
 |||
 DB 949 CTGCTGTTGAAGC 936

RESULT 43
 AAC98021/C
 ID AAC98021 standard; cDNA; 2675 BP.
 XX
 AC AAC98021;
 XX
 DT 09-MAR-2001 (first entry)
 XX
 DE Human colon cancer antigen nucleotide sequence SEQ ID NO:31.
 XX
 KW Human: colon cancer; colon cancer antigen; diagnosis; detection;
 KW identification; cytostatic; cardioactive; neuroprotective; vulnerrary;
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;
 KW nephrotic; antifungal; antibacterial; gene therapy; wound;
 KW neural disorder; immune system disorder; muscular disorder;
 KW reproductive disorder; gastrointestinal disorder; renal disorder;
 KW

```

KW infectious disease; cardiovascular disorder; ss.
XX
XX Homo sapiens.
XX
PN W0200055351-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05883.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI: 2000-587534/55.
XX
XX P-PSDB; AAB53264.
XX
XX Colon cancer associated gene sequences, referred to as colon cancer
XX PT antigens, useful for the treatment, prevention, and diagnosis of colon
XX PT disorders such as colon cancer -
XX
XX Claim 1: Page 490-491; 2104pp; English.
XX
XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
XX CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
XX CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
XX CC neuroprotective, immunomodulatory, gynecological, gastrointestinal,
XX CC vulnerrary, nephrotoxic, anti-infective and antibacterial activities, and
XX CC can be used in gene therapy. The colon cancer antigen polynucleotides,
XX CC proteins and antibodies to the proteins are useful for the prevention,
XX CC treatment and diagnosis of colon disorders, such as colon cancer. The
XX CC polynucleotides may be used in diagnostics and research, such as for
XX CC chromosome identification, and as hybridisation probes. The proteins
XX CC may also be used to prevent diseases such as neural disorders, immune
XX CC system disorders, muscular disorders, reproductive disorders,
XX CC gastrointestinal disorders, wounds, renal disorders, infectious
XX CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
XX CC AAB54007 represent sequences used in the exemplification of the present
XX CC invention.
XX
XX Sequence 2675 BP; 856 A; 425 C; 532 G; 855 T; 7 other:
XX
XX
XX Query Match          56.0%; Score 14; DB 21; Length 2675;
XX Best Local Similarity 100.0%; Pred. No. 84;
XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 11 ctctgtttgaac 24
XX      |||||
XX DB 972 CTGCTGTTGAAC 959
XX
XX RESULT 44
XX AAQ05686/C
XX ID AAQ05686 standard; DNA; 3243 BP.
XX
XX AC AAQ05686;
XX
XX DT 03-JAN-1991 (first entry)
XX
XX DE Islets of Langerhans cell clone ICA12.3 (ATCC 40703).
XX
XX KW Pancreatic Islet Cell Antigen (ICA) 12.3; ATCC 40703;
XX KW insulin-dependent (Type I) diabetes mellitus; ss;
XX
XX OS Homo, sapiens.
XX
XX FH Key Location/Qualifiers
XX FT CDS 692..1618
XX FT /*tag= a
XX

```

```

PN EP383129-A.
XX
XX 22-AUG-1990.
XX
XX 03-FEB-1990; 90EP-0102179.
XX
XX 04-DEC-1989; 89US-0441703.
XX PR 17-FEB-1989; 89US-0312543.
XX
XX (MOLE-) MOLECULAR DIAGNOSTI.
XX
XX Rabin D;
XX
XX WPI: 1990-255450/34.
XX
XX P-PSDB; AAR06467.
XX
XX Pancreatic islet cell antigens obtd. by mol. cloning - used in
XX PT research, diagnosis and treatment of insulin-dependent (Type I)
XX PT diabetes mellitus.
XX
XX Disclosure; Fig 13; 45pp; English.
XX
XX A cDNA library was constructed from RNA extracted from purified
XX CC human islets of Langerhans cells. The library was screened with
XX CC sera from autoimmune patients with Type I diabetes mellitus.
XX CC ICA 12.3 is one of nine clones found to react predominantly with
XX CC diabetic sera. The protein expressed by the clone was analysed by
XX CC expressing the clones in E.coli hosts. The protein or peptide
XX CC fragments of it can be used as immunoassay reagents in the
XX CC presymptomatic diagnosis of insulin-dependent diabetes mellitus
XX CC (IDDM).
XX CC They may also be used to identify or mark T- or B-cells involved in
XX CC IDDM or to bind or block these same cells for therapy.
XX CC See also AAQ05681-005685 and AAQ05687-005688.
XX
XX Sequence 3243 BP; 664 A; 1007 C; 853 G; 719 T; 0 other:
XX
XX
XX Query Match          56.0%; Score 14; DB 11; Length 3243;
XX Best Local Similarity 100.0%; Pred. No. 84;
XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 3 tgcctgactctg 16
XX      |||||
XX DB 413 TGCTGATCTGCTG 400
XX
XX RESULT 45
XX AA232338/C
XX ID AA232338 standard; cDNA; 3243 BP.
XX
XX AC AA232338;
XX
XX DT 21-JAN-2000 (first entry)
XX
XX DE Human pancreatic islet cell antigen ICA-12.3 ATCC 40703 cDNA sequence.
XX
XX KW Human; pancreatic islet cell antigen; ICA clone; IDDM; diabetes;
XX KW insulin dependent type I diabetes mellitus; diagnosis; detection;
XX KW immunoglobulin; T-cell; B-cell; antibody binding; ss.
XX
XX OS Homo sapiens.
XX
XX PN US5981700-A.
XX
XX PD 09-NOV-1999.
XX
XX PF 06-JUN-1995; 95US-0468579.
XX
XX PR 05-MAY-1994; 94US-0239276.
XX PR 14-JUN-1991; 91US-0715181.
XX PR 08-JUN-1992; 92US-0872646.
XX PR 17-FEB-1989; 89US-0312543.
XX

```

```

PR 04-DEC-1989; 89US-041703.
XX
XX (FARB ) BAYER CORP FORMERLY MOLECULAR DIAGNOSTIC.
XX
XX Rabin DU:
PI
XX WPI: 1999-633370/54.
DR P-PDB: AAY49852.
XX
XX New pancreatic islet cell antigens, useful in the diagnosis of
PT insulin-dependent (Type 1) diabetes mellitus.
XX
XX Claim 1: Column 35-42; 61pp; English.
XX
XX The present sequence encodes a human pancreatic islet cell antigen (ICA)
CC that binds with antibodies found in the sera of patients afflicted with
CC insulin-dependent (Type 1) diabetes mellitus (IDDM). ICAs from the
CC present invention are encoded by the DNA insert of a recombinant cloning
CC vehicle selected from ATCC 40550 (AA232333, encoding AAY49847), ATCC
CC 40553 (AA232334, encoding AAY49848), ATCC 40554 (AA232335, encoding
CC AAY49849), ATCC 40551 (AA232336, encoding AAY49850), ATCC 40552
CC (AA232337, encoding AAY49851), ATCC 40703 (AA232338, encoding AAY49852),
CC ATCC 40704 (AA232339, encoding AAY49853), ATCC 40705 (AA232340), ATCC
CC 40706 (AA232341, encoding AAY49854) and ATCC 75030 (AA232342, encoding
CC AAY49855). ICA proteins and their peptide fragments can be used in the
CC diagnosis of IDDM and in detecting or blocking human immunoglobulin,
CC T-cells or B-cells involved in IDDM.
XX
XX Sequence 3243 BP; 664 A; 1007 C; 853 G; 719 T; 0 other;
SQ

```

```

Query Match          56.0%; Score 14; DB 20; Length 3243;
Best local Similarity 100.0%; Pred No. 84;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 3 tgcctgagatctgctg 16
   |||
Db 413 TGCTGGATCTGCTG 400

```

Search completed: October 9, 2001, 15:52:08
 Job time: 15089 sec

